

<http://www.diva-portal.org>

This is the published version of a paper published in *Journal of Mathematical Imaging and Vision*.

Citation for the original published paper (version of record):

Lindeberg, T., Babaiee, Z., Kiasari, P M. (2026)

Modeling and analysis of the 8 filters from the “master key filters hypothesis” for depthwise-separable deep networks in relation to idealized receptive fields based on scale-space theory

*Journal of Mathematical Imaging and Vision*, 68(3): 22:1-22:26

<https://doi.org/10.1007/s10851-026-01290-0>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-382525>



# Modeling and Analysis of the 8 Filters from the “Master Key Filters Hypothesis” for Depthwise-Separable Deep Networks in Relation to Idealized Receptive Fields Based on Scale-Space Theory

Tony Lindeberg<sup>1</sup> · Zahra Babaiee<sup>2</sup> · Peyman M. Kiasari<sup>2</sup>

Received: 16 September 2025 / Accepted: 9 March 2026  
© The Author(s) 2026

## Abstract

This paper presents the results of analyzing and modeling a set of 8 “master key filters”, which have been extracted by applying a clustering approach to the receptive fields learned in depthwise-separable deep networks based on the ConvNeXt architecture. For this purpose, we first compute spatial spread measures in terms of weighted mean values and weighted variances of the absolute values of the learned filters, which support the working hypotheses that: (i) the learned filters can be modeled by separable filtering operations over the spatial domain, and that (ii) the spatial offsets of the those learned filters that are non-centered are rather close to half a grid unit. Then, we model the clustered “master key filters” in terms of difference operators applied to a spatial smoothing operation in terms of the discrete analog of the Gaussian kernel, and demonstrate that the resulting idealized models of the receptive fields show good qualitative similarity to the learned filters. This modeling is performed in two different ways: (i) using possibly different values of the scale parameters in the coordinate directions for each filter, and (ii) using the same value of the scale parameter in both coordinate directions. Then, we perform the actual model fitting by either (i) requiring spatial spread measures in terms of spatial variances of the absolute values of the receptive fields to be equal, or (ii) minimizing the discrete  $l_1$ - or  $l_2$ -norms between the idealized receptive field models and the learned filters. Complementary experimental results then demonstrate that the idealized models of receptive fields have very good predictive properties for replacing the learned filters by idealized filters in depthwise-separable deep networks, thus showing that the learned filters in depthwise-separable deep networks can be well approximated by discrete scale-space filters. Notably, we show that, for a reduced version of the ConvNeXt architecture, using a set of only 8 discrete scale-space filters leads to almost as good accuracy as for the receptive fields trained from scratch on the ImageNet dataset.

**Keywords** Receptive field · Deep learning · Discrete · Continuous · Gaussian kernel · Gaussian derivative · Depthwise-separable networks · Scale space

## 1 Introduction

In computer vision operations, specifically such operations formulated in terms of deep networks applied to image data, a main computational function is performed in terms of the

receptive fields, that integrate the image information over non-infinitesimal support regions over the image domain. A both theoretical and practical problem does hence concern how to choose the models to use for such receptive fields in computer vision algorithms.

In the area of scale-space theory, the problem of choosing appropriate models for receptive fields has been addressed from a normative theoretical viewpoint, by formulating assumptions, referred to as scale-space axioms, that reflect desirable properties of a vision system under selected families of symmetry transformations. By axiomatic derivations, it has in this respect been shown that Gaussian kernels and Gaussian derivatives constitute a canonical family of linear filters to be used in the first layer of the image operations, see Iijima [12], Koenderink [14], Koenderink and van Doorn

---

The support from the Swedish Research Council (contract 2022-02969) is gratefully acknowledged.

---

✉ Tony Lindeberg  
tony@kth.se

<sup>1</sup> Computational Brain Science Lab, Department of Computational Science and Technology, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

<sup>2</sup> TU Wien, Informatics, 1040 Vienna, Austria

[15], Lindeberg [20, 21, 23] and Weickert et al. [36]. The idealized models for spatial receptive fields obtained in this way have also been shown to agree qualitatively quite well with biological receptive fields registered by neurophysiological recordings in the retina, the lateral geniculate nucleus (LGN) and the primary visual cortex (V1) of higher mammals, see Lindeberg [22, 23].

In the area of deep learning, the problem of choosing the receptive fields is on the other hand addressed from a pure learning perspective, where the resulting receptive fields then arise as the result of optimizing some suitably selected loss function for a specific deep learning architecture. Hybrid approaches to this problem have also been formulated, where the use of Gaussian derivative operators has been extended to higher layers in deep networks by Jacobsen et al. [13], Lindeberg [24, 25], Pintea et al. [32], Sangalli et al. [34], Penaud-Polge et al. [30], Smets et al. [35], Gavilima-Pilataxi and Ibarra-Fiallo [9] and Perzanowski and Lindeberg [31]. The use of Gaussian derivative operators in the higher layers of such hybrid networks has, however, not previously been theoretically motivated by necessity, as the use of Gaussian derivatives in the first layer of a vision system has been formally established, based on the axiomatically derived scale-space theory by Iijima [12], Koenderink [14], Koenderink and van Doorn [15], Lindeberg [20, 21, 23] and Weickert et al. [36].

More recently, Babaiee et al. [2–4] have investigated the properties of depthwise convolutional filters learned across all the layers of depthwise-separable convolutional neural networks, as originally proposed by Chollet [7] and Howard et al. [11], with particular focus on the ConvNeXt architectures by Liu et al. [29] and Woo et al. [37]. Through unsupervised clustering techniques applied to millions of trained filters, they discovered that the depthwise convolutional kernels can be categorized into a few distinct classes, resembling Gaussian functions and their derivatives. Remarkably, these filter patterns demonstrate generality regardless of the layer depth, the network architecture, or the training data domain. Through systematic greedy search methods, they further demonstrated that this diverse set of learned filters can, without major loss of performance, be reduced to a smaller set of just 8 so-called “master key filters”, see Fig. 1 and the first row in Fig. 2 for two different ways of visualizing the same data.<sup>1</sup>

<sup>1</sup> Concerning the visualizations of these filters, let us remark that that the actual scaling or the base levels of these filters is not regarded as essential for the functionality of the resulting deep networks in this study, since (i) the scaling of the filters can both be propagated across the layers in the network hierarchy with maintained essential functionality regarding the computations and can also be compensated for by rescalings of the weights that combine the output from the different filters, and (ii) the base levels of the filters can be compensated for with the bias terms in the linear combination of filter outputs before the nonlinearity in

Notably, these receptive fields do specifically appear to be qualitatively very similar to receptive fields obtained from discrete scale-space theory, in terms of either (see rows 2–7 in Fig. 2 for illustrations):

- anti-symmetric first-order derivative approximations at moderately fine scales (Filters 5 and 6),
- a Gaussian blob at a moderately fine scale (Filter 8),
- a local sharpening operation by subtracting a Laplacian-of-the-Gaussian at a moderately fine scale (Filter 7), or
- non-centered asymmetric first-order derivative approximations at very fine scales (Filters 1–4).

The subject of this paper is to present a quantitative analysis of these learned filters, and specifically to make comparisons to idealized simplifications with different degrees of simplification in idealized models for spatial receptive fields, based on an earlier presented theory for discrete derivative approximations with scale-space properties in Lindeberg [18, 19, 27, 28].

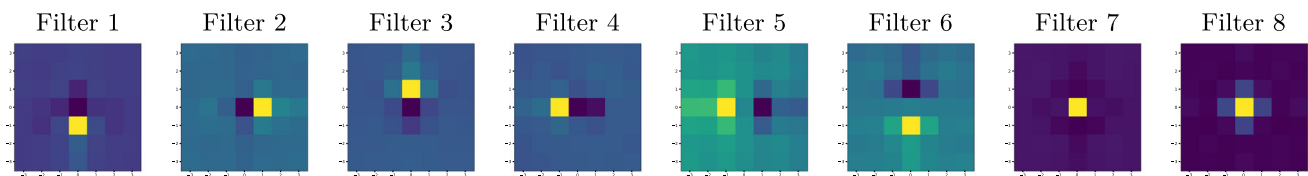
In summary, the results to be presented will show that the learned filters in depthwise-separable deep networks based on the ConvNeXt architecture can be well modeled by discrete scale-space filters. By quantitative results regarding deep learning, we also show that replacement of the learned filter by idealized scale-space filters in depthwise-separable deep networks also lead to almost as good results as when using the fully learned filters. In these respects, the results presented in this paper generalize the previous solid theoretical foundation of using Gaussian derivative filters as the first layer of linear filters in a vision system into a combined theoretical and experimentally based foundation to also using (here discrete approximations of) Gaussian derivative operators in the higher layers of deep networks.

Given that the ConvNeXt architecture has been demonstrated to constitute a modern state-of-the-art architecture for deep networks with very good ability to compete with Transformer architectures, a main significance of this work is that it establishes a strong link between the filters learned in a reduced ConvNeXt architecture and the idealized filters predicted from theoretically based scale-space theory.

## 1.1 Contributions and Novelty

In brief, this paper comprises the following theoretical, conceptual and experimental contributions:

each computational unit. Instead, for our approach, the spatial variations of the filters are regarded as providing the essence of the underlying computations, since these filter shapes determine how the local image information is to be combined between different spatial points in the image domain.



**Fig. 1** Original set of 8 “master key filters”, obtained by applying a clustering technique to the receptive fields learned from depthwise-separable deep networks, as extracted by Babaiee et al. [4], “through greedy search on the ConvNeXt V2 Tiny model” developed by Woo et

al. [37], in turn based on the regular ConvNeXt model developed by Liu et al. [29]. (Horizontal axes: horizontal filter indices  $m \in [-3, 3]$ . Vertical axes: vertical filter indices  $n \in [-3, 3]$ .)

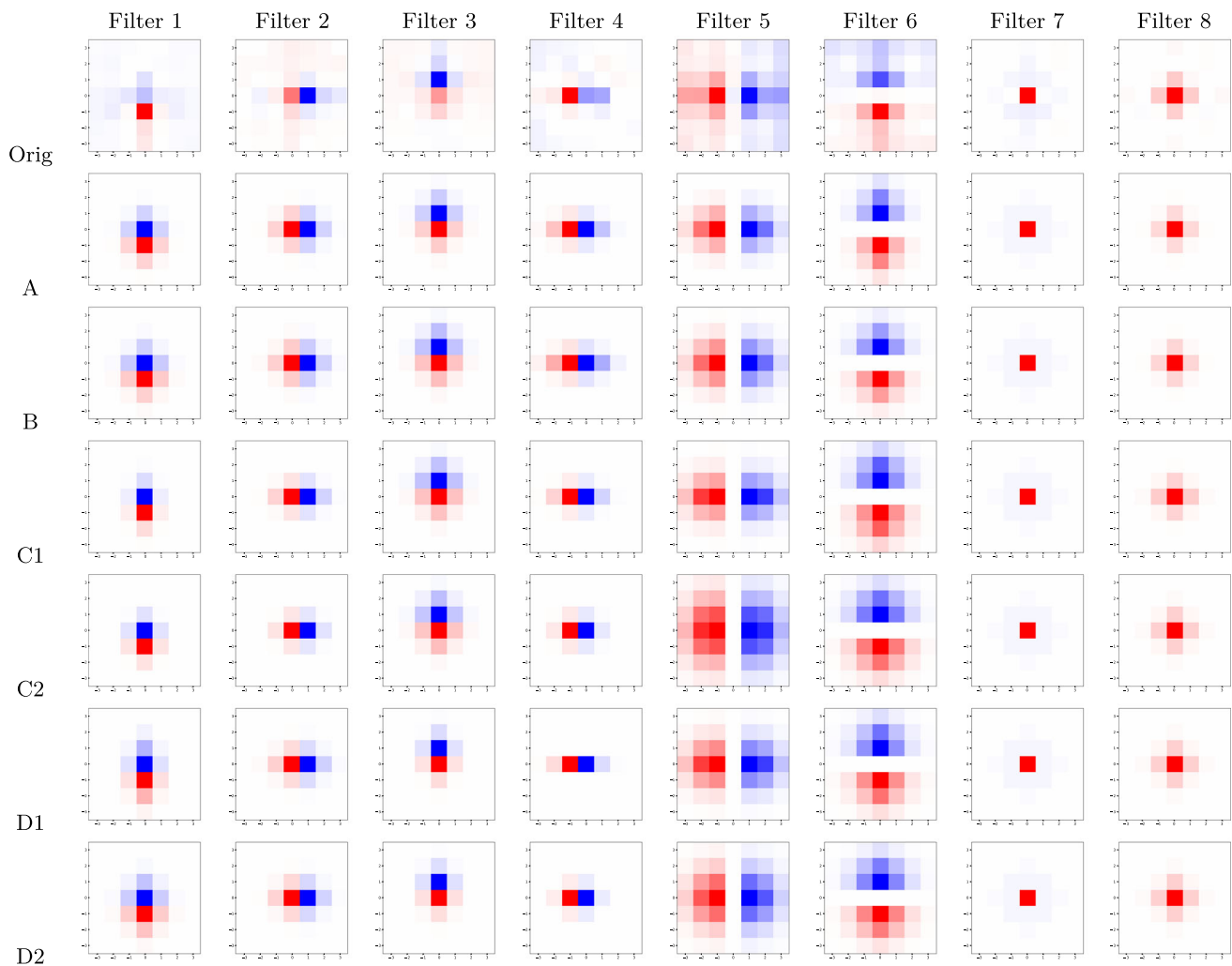
- Extensions of the discrete scale-space theory to using non-centered discrete filter models.
- A methodology for characterizing the properties of both idealized receptive field models and learned filters based on scale-space theory.
- Methods for estimating the scale parameters in idealized receptive field models from learned filters based on matching of either continuous or discrete mean-based and variance-based spatial spread measures.
- The formulation of weighted spatial spread measures, to reduce the risk that contributions from spurious values in the background may either lead to overestimates (biases) in the scale parameters.
- The set of 8 idealized receptive field models obtained by model fitting for the 8 “master key filters”.
- The proposal of replacing the learned filters in the ConvNeXt V2 architecture with idealized receptive fields based on scale-space theory, and evaluating the accuracy obtained when using different types of model fitting for determining the scale parameters in the idealized models of the receptive fields.
- The result that model fitting based on discrete weighted spatial spread measures leads to significantly better predictive properties regarding the performance of using idealized receptive fields in a deep network architecture, compared to either using weighted continuous spatial spread measures or more traditional  $l_1$ - or  $l_2$ -based model fitting.
- The result that using idealized receptive fields obtained from scale-space theory leads to almost as good accuracy as when using the originally learned filters in the ConvNeXt V2 Tiny architecture.
- The result that only a marginal increase in the accuracy is obtained by replacing the values of the scale parameters obtained applying the best Method B to the 8 “master key filters” with the values of the scale parameters obtained by data-driven learning of the scale parameters using back-propagation.
- A theoretically based interpretation of the spaces spanned by different sets of discrete derivative approximation operators applied to discrete Gaussian smoothing filters.
- A set of predictions regarding the choices of the sets of filters to be used in further work regarding deep networks based on Gaussian derivatives or other closely related scale-space filters.

In these ways, the results provide additional support for using idealized scale-space filters as the receptive field primitives in deep networks, thereby extending (i) the previous axiomatic *necessity* results of using scale-space filters in the first layer in the visual hierarchy, as well as (ii) the previous *sufficiency* results of using Gaussian derivative kernels in higher layers of deep networks, toward (iii) showing that the receptive fields learned in all the layers of a modern state-of-the-art architecture for deep learning can be *very well approximated* by idealized scale-space filters.

## 1.2 Structure of this Paper

The presentation is organized as follows: Sect. 2 provides a conceptual background regarding the ConvNeXt architecture for deep learning and the procedure for extracting the set of 8 “master key filters” from the filters learned from this architecture. Sect. 3 then starts by describing a theoretical background for measuring characteristic properties of the learned filters, based on spatial spread measures, defined from spatial means and spatial variances of the absolute values of the filter coefficients, as well as the responses to lower-order polynomials of the learned filters. Section 4 continues by performing modeling of the 8 “master key filters” using four different types of main methods in terms of:

- Method A: Spatial spread measures as matched between discrete spatial spread measures of the learned filters and continuous spatial spread measures of continuous Gaussian derivatives.
- Method B: Spatial spread measures as matched between discrete spatial spread measures of the learned filters and discrete spatial spread measures of discrete scale-space kernels defined by applying small support difference operators to the discrete analog of the Gaussian kernel.



**Fig. 2** Visualizations of the learned filters with the corresponding results of fitting idealized models of these filters using scale-space operations: **(top row)** Alternative visualization of the original set of 8 “master key filters”, extracted by Babaie et al. [4], while here complemented with a normalization the filters, (i) by rescaling Filters 1–6 to give the same response to matching first-order discrete monomials under discrete convolution as for convolution of the corresponding continuous monomials with corresponding continuous Gaussian derivative operators, according to (63)–(68), and (ii) by adding different constants to Filters 7–8, according to (69)–(73), to minimize the variance-based spatial spread measure of the filters, and finally visualizing the data on a blue-red color scale with the value 0 corresponding to white, to better reveal the polarities of the filter values. **(rows 2–7)** Idealized scale-space models of the filters, as computed with the different types of modeling approaches proposed in this paper: **Method A** in Sect. 4.1, based on direct transfer of scale values from the variances for a continuous Gaussian derivative model, **Method B** in Sect. 4.2, based on requiring the horizontal vs. the vertical discrete weighted variance-based spatial spread measures for the idealized receptive field models to be equal to the weighted discrete spatial spread measures for the corresponding

learned filters, **Method C1** in Sect. 4.3, based on minimizing the discrete  $l_1$ -norm between the idealized receptive field models and the normalized versions of the learned filters and using different values of the scale parameters in the horizontal and the vertical directions, **Method C2** in Sect. 4.4, based on minimizing the discrete  $l_1$ -norm between the idealized receptive field models and the normalized versions of the learned filters and using the same values of the scale parameters in the horizontal and the vertical directions, **Method D1** in Sect. 4.5, based on minimizing the discrete  $l_2$ -norm between the idealized receptive field models and the normalized versions of the learned filters and using different values of the scale parameters in the horizontal and the vertical directions, and **Method D2** in Sect. 4.6, based on minimizing the discrete  $l_2$ -norm between the idealized receptive field models and the normalized versions of the learned filters and using the same values of the scale parameters in the horizontal and the vertical directions. (Note that the contrasts of Filters 2 and 3 are reversed in relation to the sign conventions for the corresponding “master key filters”.) (Horizontal axes: horizontal filter indices  $m \in [-3, 3]$ . Vertical axes: vertical filter indices  $n \in [-3, 3]$ .)

- Method C: Minimization of the difference between the learned filters and idealized discrete scale-space filters in discrete  $l_1$ -norm.
- Method D: Minimization of the difference between the learned filters and idealized discrete scale-space filters in discrete  $l_2$ -norm.

Section 5 then interprets the results obtained in this way in terms of the spaces spanned by the receptive field responses in terms of  $N$ -jet representations based on Gaussian derivatives. This section also describes implications and predictions from these results regarding the construction of more general Gaussian derivative networks, with the layers defined from linear combinations of Gaussian derivative responses.

Section 6 thereafter applies the fitted idealized discrete scale-space filters as the filtering primitives in depthwise-separable networks based on the ConvNeXt V2 Tiny architecture applied to the ImageNet dataset, and shows that:

- the idealized filters based on Method B have the best predictive properties out of the four main classes of filter modeling methods, and
- using the idealized models of the 8 “master key filters” as the computational primitives in depthwise-separable deep networks leads to almost as good results as if using the originally learned filters.

Finally, Sect. 7 concludes with a summary and discussion of some of the main results.

## 2 Background on Depthwise-Separable CNNs and the ConvNeXt Architecture

### 2.1 Convolutional Neural Networks and Depthwise-Separable Convolutions

Convolutional Neural Networks (CNNs) constitute a class of deep learning architectures in which spatial filtering operations are performed by learned convolutional kernels applied across multiple layers [16]. In a standard convolutional layer, each output channel is obtained by convolving all input channels with a kernel tensor of size  $k \times k \times C_{\text{in}}$ , where  $k$  denotes the spatial kernel size and  $C_{\text{in}}$  the number of input channels. Thus, spatial filtering and cross-channel mixing are performed jointly.

Depthwise-separable convolutions decouple these two operations into a spatial and a channel mixing stage [7, 11]. Specifically:

- **Depthwise convolution:** A spatial convolution with a  $k \times k$  kernel is applied independently to each input channel. If

the input has  $C$  channels, this stage uses  $C$  spatial kernels and does not mix information across channels.

- **Pointwise convolution:** A so-called  $1 \times 1$  convolution is subsequently applied to linearly combine information across channels.

If both the input and output have  $C$  channels, a standard convolution requires  $k^2 C^2$  parameters, whereas a depthwise-separable convolution requires  $k^2 C + C^2$  parameters. For moderate  $k$  and large  $C$ , this leads to substantial computational savings.

From a structural viewpoint, the depthwise stage performs the entire spatial filtering operation. The subsequent pointwise convolution operates purely over the channel dimension.

### 2.2 The ConvNeXt Architecture

ConvNeXt is a modern convolutional architecture designed to achieve competitive performance with Transformer-based vision models, while retaining a fully convolutional structure [29]. It can be interpreted as a modernization of classical residual networks [10], incorporating architectural refinements inspired by Vision Transformers [8].

A ConvNeXt block typically consists of:

1. a depthwise convolution (often with kernel size  $7 \times 7$ ),
2. a layer normalization (LN<sup>2</sup>),
3. a pointwise ( $1 \times 1$ ) convolution expanding the channel dimension,
4. a nonlinearity (e.g., GELU<sup>3</sup>),
5. a global response normalization (GRN<sup>4</sup>) layer, which enhances the feature diversity, by encouraging competition between the channels [37],
6. a second pointwise convolution, projecting back to the original channel dimension, and
7. a residual connection.

Crucially, the depthwise convolution is the only operation within the block that performs spatial filtering. All the sub-

<sup>2</sup> Instead of normalizing across the batch, like standard CNNs, the LayerNorm (LN) operation in the ConvNeXt V2 architecture normalizes across the channels for each individual pixel (or spatial location), helping to stabilize the training and mimicking the behavior of Vision Transformers.

<sup>3</sup> The Gaussian Error Linear Unit (GELU) is a smoother non-monotonic replacement of the ReLU function,  $\text{GELU}(x) = \frac{x}{2} \left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$ , found to enhance the training in deep networks.

<sup>4</sup> Introduced in ConvNeXt V2, the global response normalization (GRN) layer enhances the feature diversity, by normalizing the activations of the channels based on their global relative importance, thereby preventing any single “dead” or over-dominant channel from stalling the network.

sequent operations act pointwise across spatial locations. Therefore, the spatial receptive field structure learned by ConvNeXt is entirely determined by its depthwise kernels.

The global architecture of the ConvNeXt V2 network, that we will build our work upon, is organized into four hierarchical stages with a block distribution of (3, 3, 9, 3) for the Tiny variant, as illustrated in Fig. 3. Starting from an initial patching layer with  $C = 96$  channels, with convolutional kernels and a stride size of 4, the channel dimensions double at each stage (96, 192, 384, 768), using strided convolutions for downsampling.

### 2.3 The Master Key Filters Hypothesis

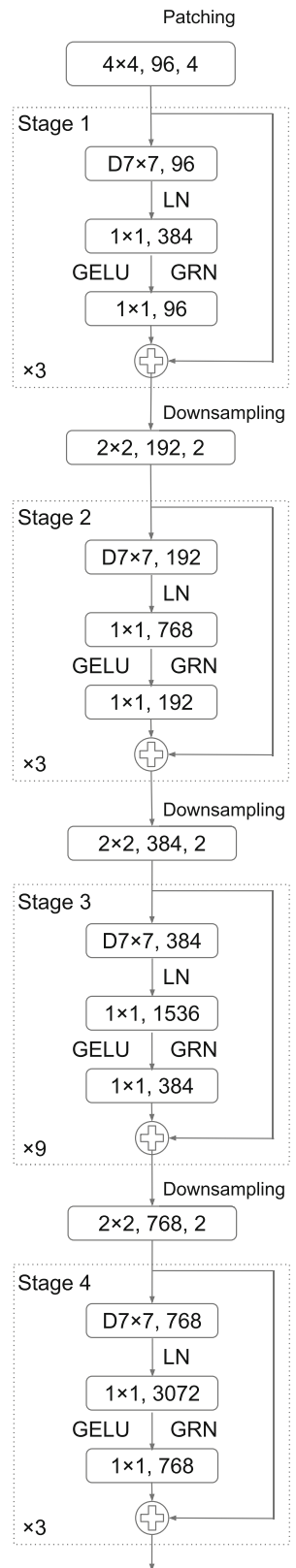
Research shows that the trained filters in depthwise-separable CNNs consistently converge into a few identifiable clusters across all the layers [3]. Through unsupervised clustering of millions of filters, these patterns were found to resemble Gaussian functions and derivatives of Gaussian kernels. This discovery led to the Master Key Filters Hypothesis, which proposes that a general set of universal filters exists for visual data and that depthwise filters in depthwise-separable CNNs naturally tend toward these “master keys” regardless of the specific dataset, architecture, or layer depth [4].

To find the minimal set of the master key filters, an autoencoder was used to compress the trained  $7 \times 7$  depthwise kernels into a single-dimensional hidden code [5]. By sampling points from this latent 1D space, a spectrum of candidate filters was constructed. To find the minimal essential set, a greedy search was performed on the filters learned in the ConvNeXt V2 Tiny architecture, replacing its original filters with the closest linear transformations of these sampled candidates. The least important ones were removed iteratively, by measuring the change in accuracy, when each candidate was removed from the list. This process identified a distinct “elbow point” on the accuracy plot at precisely 8 filters, where the accuracy remained stable up to that point despite a massive reduction in unique weights. These 8 essential filters, shown in Fig. 1, can effectively replace thousands of trained parameters, while maintaining comparable performance on ImageNet.

### 2.4 Tangible Contributions to ConvNeXt Networks

In this work, we start from the given set of 8 “master key filters”, extracted using the procedure described above, and model these filters in terms of idealized models of receptive fields, as obtained from scale-space theory. By measuring and characterizing the spatial extents of these filters in terms of spatial spread measures, we show that the 8 “master key filters” can be well modeled by a set of 8 discrete scale-space filters.

**Fig. 3** Architectural overview of the ConvNeXt V2 Tiny network. Here, the abbreviation “LN” denotes “Layer Normalization”, while “GRN” denotes “Global Response Normalization”



By then replacing the originally learned filters in the ConvNeXt V2 Tiny architecture with these idealized models of spatial receptive fields, we show that it is possible to obtain almost as good results on the ImageNet dataset as if using the originally learned filters. This result thereby shows that the filters in the ConvNeXt V2 Tiny architecture can be very well approximated by discrete scale-space filters.

### 3 Measuring Characteristic Properties of the 8 “Master Key Filters” in Terms of Spatial Spread Measures and Responses to Lower-Order Polynomials

In this section, we will apply a similar methodology, as used for modeling and characterizing the discrete derivative approximation methods in Lindeberg [27, 28], to model the 8 “master key filters” extracted by Babaiee et al. [4], while also employing structurally closely related methods for determining the parameters of idealized receptive field models, as previously used for matching the parameters between different types of theoretical models for temporal receptive fields in Lindeberg ([26], Sections 3.3–3.4).

#### 3.1 Spatial Spread Measures for Characterizing the Spatial Extent and the Spatial Offset of Receptive Fields

For measuring the overall effect of a non-negative receptive field  $h(p) = h(x, y)$  with the image coordinates  $p = (x, y)^T$  with regard to its spatial extent, it is natural to formulate a spatial spread measure in terms of the spatial mean  $M(h(\cdot))$  and the spatial variance  $V(h(\cdot, \cdot))$  according to

$$M(h(\cdot)) = \frac{\int_{p \in \mathbb{R}^2} p h(p) dp}{\int_{p \in \mathbb{R}^2} h(p) dp}, \tag{1}$$

$$V(h(\cdot)) = \frac{\int_{p \in \mathbb{R}^2} p p^T h(p) dp}{\int_{p \in \mathbb{R}^2} h(p) dp} - \left( \frac{\int_{p \in \mathbb{R}^2} p h(p) dp}{\int_{p \in \mathbb{R}^2} h(p) dp} \right)^2, \tag{2}$$

over a continuous image domain and according to

$$M(h(\cdot)) = \frac{\sum_{p \in \mathbb{R}^2} p h(p)}{\sum_{p \in \mathbb{R}^2} h(p)}, \tag{3}$$

$$V(h(\cdot)) = \frac{\sum_{p \in \mathbb{R}^2} p p^T h(p)}{\sum_{p \in \mathbb{R}^2} h(p)} - \left( \frac{\sum_{p \in \mathbb{R}^2} p h(p)}{\sum_{p \in \mathbb{R}^2} h(p)} \right)^2, \tag{4}$$

over a discrete image domain.

The use of these spatial spread measures is structurally related to the use of mean values and covariance matrices for characterizing gross overall properties of statistical distributions. For our purpose, of capturing essential properties of both the learned filters in a deep network architecture and the use of idealized filter models to replace those filters, the motivation behind using these measures is to in a compact manner characterize how both the learned filters and the idealized filters weigh the image information from different spatial points.

In previous work, the use of such spatial spread measure has been demonstrated to have a predictive ability, to compare the effective amount of spatial smoothing in receptive fields, when performing learning of the scale levels in Gaussian derivative networks using different types of discretization operations for Gaussian derivative operators, see Perzanowski and Lindeberg [31].

##### 3.1.1 Spatial Spread Measures Applied to Zero-Order Scale-Space Kernels

For the continuous 2-D Gaussian kernel

$$g(x, y; \sigma) = g_{1D}(x; \sigma) g_{1D}(y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}, \tag{5}$$

with the 1-D Gaussian kernel given by

$$g_{1D}(x; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \tag{6}$$

with standard deviation  $\sigma$  and the spatial variance  $s = \sigma^2$ , the resulting variance-based spread measure is then

$$V(g(\cdot, \cdot; \sigma)) = \sigma^2 I, \tag{7}$$

where  $I$  denotes the unit matrix. Similarly, for the 2-D discrete analog of the Gaussian kernel [17, 27]

$$T(m, n; \sigma) = T_{1D}(m; \sigma) T_{1D}(n; \sigma) \tag{8}$$

of the Gaussian kernel given by

$$T_{1D}(n; \sigma) = e^{-\sigma^2} I_n(\sigma^2), \tag{9}$$

where  $I_n(\sigma^2)$  denotes the modified Bessel functions of integer order (see Abramowitz and Stegun [1]), the corresponding discrete spatial spread measure is correspondingly [18, 27]

$$V(T(\cdot, \cdot; \sigma)) = \sigma^2 I. \tag{10}$$

Since both the continuous Gaussian kernel and the discrete analog of the Gaussian kernel are centered, we also have that their spatial mean values are

$$M(g(\cdot, \cdot; \sigma)) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{11}$$

$$M(T(\cdot, \cdot; \sigma)) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{12}$$

### 3.1.2 Spatial Spread Measures Applied to First-Order Scale-Space Kernels

Following the methodology in Lindeberg [27, 28], we will in this paper also perform corresponding characterizations of receptive fields  $h(x, y)$  that are not guaranteed to be non-negative, by then measuring the mean value and the variance of the absolute value of the receptive field  $|h(x, y)|$ , in terms of the entities  $M(|h(\cdot, \cdot)|)$  and  $V(|h(\cdot, \cdot)|)$ .

By computing these spatial spread measures for the horizontal first-order Gaussian derivative kernel

$$g_x(x, y; \sigma) = \partial_x(g(x, y; \sigma)) = -\frac{x}{\sigma^2} g(x, y; \sigma), \tag{13}$$

we obtain

$$M(|g_x(\cdot, \cdot; \sigma)|) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{14}$$

$$V(|g_x(\cdot, \cdot; \sigma)|) = \begin{pmatrix} 2\sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}. \tag{15}$$

Similarly, for the vertical first-order Gaussian derivative kernel

$$g_y(x, y; \sigma) = \partial_y(g(x, y; \sigma)) = -\frac{y}{\sigma^2} g(x, y; \sigma), \tag{16}$$

it holds that

$$M(|g_y(\cdot, \cdot; \sigma)|) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{17}$$

$$V(|g_y(\cdot, \cdot; \sigma)|) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix}. \tag{18}$$

Later, we will make use of these results in combination with computed discrete variances of discrete kernels, that resemble first-order derivative approximations, for estimating the scale parameter  $\sigma$  in corresponding idealized models of receptive fields in terms of scale-space operations.

### 3.1.3 Spatial Spread Measures Applied to Zero-Order Pure Difference Operators

In this paper, we will in the modeling and the analysis stages primarily consider the following non-centered first-order dif-

ference operators (see Fig. 4 for an illustration):

$$(\delta_{x+h})(m, n) = h(m + 1, n) - h(m, n), \tag{19}$$

$$(\delta_{x-h})(m, n) = h(m, n) - h(m - 1, n), \tag{20}$$

$$(\delta_{y+h})(m, n) = h(m, n + 1) - h(m, n), \tag{21}$$

$$(\delta_{y-h})(m, n) = h(m, n) - h(m, n - 1), \tag{22}$$

and the following centered first-order difference operators:

$$(\delta_x h)(m, n) = (h(m + 1, n) - h(m - 1, n))/2, \tag{23}$$

$$(\delta_y h)(m, n) = (h(m, n + 1) - h(m, n - 1))/2. \tag{24}$$

Additionally, in the following interpretations of the results to be considered in Sect. 5, we will also consider the following centered second-order difference operators:

$$(\delta_{xx} h)(m, n) = h(m + 1, n) - 2h(m, n) + h(m - 1, n), \tag{25}$$

$$(\delta_{xy} h)(m, n) = (h(m + 1, n + 1) - h(m + 1, n - 1) - h(m - 1, n + 1) + h(m - 1, n - 1))/4, \tag{26}$$

$$(\delta_{yy} h)(m, n) = h(m, n + 1) - 2h(m, n) + h(m, n - 1). \tag{27}$$

For the non-centered first-order difference operators, we specifically have the following concerning their spatial mean values:

$$M(|\delta_{x+}|) = \begin{pmatrix} +\frac{1}{2} \\ 0 \end{pmatrix}, \tag{28}$$

$$M(|\delta_{x-}|) = \begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}, \tag{29}$$

$$M(|\delta_{y+}|) = \begin{pmatrix} 0 \\ +\frac{1}{2} \end{pmatrix}, \tag{30}$$

$$M(|\delta_{y-}|) = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}, \tag{31}$$

while for the centered first-order or second-order difference operators,  $\delta_* \in \{\delta_x, \delta_y, \delta_{xx}, \delta_{xy}, \delta_{yy}\}$ , because of their pure mirror symmetries or anti-symmetries, their spatial mean values are all zero:

$$M(|\delta_*|) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{32}$$

For the non-centered first-order difference operators, we specifically have the following concerning their spatial variances:

$$V(|\delta_{x+}|) = V(|\delta_{x-}|) = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{pmatrix}, \tag{33}$$

$$\begin{aligned} \delta_{x+} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & +1 \\ 0 & 0 & 0 \end{pmatrix} & \delta_{x-} &= \begin{pmatrix} 0 & 0 & 0 \\ -1 & +1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \delta_{y+} &= \begin{pmatrix} 0 & +1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \delta_{y-} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & +1 & 0 \\ 0 & -1 & 0 \end{pmatrix} \\ \delta_x &= \begin{pmatrix} 0 & 0 & 0 \\ -1/2 & 0 & +1/2 \\ 0 & 0 & 0 \end{pmatrix} & \delta_y &= \begin{pmatrix} 0 & +1/2 & 0 \\ 0 & 0 & 0 \\ 0 & -1/2 & 0 \end{pmatrix} & \delta_{xx} &= \begin{pmatrix} 0 & 0 & 0 \\ +1 & -2 & +1 \\ 0 & 0 & 0 \end{pmatrix} & \delta_{xy} &= \begin{pmatrix} -1/4 & 0 & +1/4 \\ 0 & 0 & 0 \\ +1/4 & 0 & -1/4 \end{pmatrix} & \delta_{yy} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{aligned}$$

**Fig. 4** Computational molecules visualizing the computational functions of the different (top row) non-centered first-order difference operators and (bottom row) centered first-order and second-order difference operators considered in this paper

$$V(|\delta_{y+}|) = V(|\delta_{y-}|) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \tag{34}$$

the following concerning the centered first-order difference operators:

$$V(|\delta_x|) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{35}$$

$$V(|\delta_y|) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{36}$$

and the following concerning the centered second-order difference operators:

$$V(|\delta_{xx}|) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}, \tag{37}$$

$$V(|\delta_{xy}|) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{38}$$

$$V(|\delta_{yy}|) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{39}$$

**Table 1** Result of computing the discrete  $l_1$ -norm  $\|h(\cdot, \cdot)\|_1$  as well as the discrete spatial mean value  $M(|h(\cdot, \cdot)|)$  according to Eq. (3) and the discrete spatial variance  $V(|h(\cdot, \cdot)|)$  according to Eq. (4) for the 8 “master key filters” shown in Fig. 1

Filter	$\ h_i(\cdot, \cdot)\ _1$	$M( h_i(\cdot, \cdot) )$	$V( h_i(\cdot, \cdot) )$
1	2.47	$\begin{pmatrix} -0.023 \\ -0.602 \end{pmatrix}$	$\begin{pmatrix} +1.799 & -0.022 \\ -0.022 & +1.632 \end{pmatrix}$
2	2.56	$\begin{pmatrix} +0.559 \\ +0.092 \end{pmatrix}$	$\begin{pmatrix} +1.328 & +0.012 \\ +0.012 & +1.425 \end{pmatrix}$
3	2.73	$\begin{pmatrix} +0.056 \\ +0.693 \end{pmatrix}$	$\begin{pmatrix} +1.584 & +0.035 \\ +0.035 & +1.364 \end{pmatrix}$
4	2.32	$\begin{pmatrix} -0.579 \\ +0.098 \end{pmatrix}$	$\begin{pmatrix} +1.635 & +0.051 \\ +0.051 & +1.253 \end{pmatrix}$
5	4.44	$\begin{pmatrix} +0.057 \\ -0.041 \end{pmatrix}$	$\begin{pmatrix} +3.847 & -0.011 \\ -0.011 & +1.848 \end{pmatrix}$
6	3.81	$\begin{pmatrix} -0.006 \\ +0.086 \end{pmatrix}$	$\begin{pmatrix} +1.634 & -0.080 \\ -0.080 & +3.347 \end{pmatrix}$
7	1.96	$\begin{pmatrix} +0.043 \\ -0.086 \end{pmatrix}$	$\begin{pmatrix} +1.276 & +0.003 \\ +0.003 & +1.547 \end{pmatrix}$
8	3.07	$\begin{pmatrix} -0.029 \\ -0.048 \end{pmatrix}$	$\begin{pmatrix} +2.381 & -0.020 \\ -0.020 & +2.356 \end{pmatrix}$

### 3.2 Measuring the Spatial Extent and the Spatial Offset of the 8 “Master Key Filters”

Table 1 shows the result of computing the discrete  $l_1$ -norm

$$\|h(\cdot, \cdot)\|_1 = \sum_{(m,n) \in \mathbb{Z}^2} |h(m, n)| \tag{40}$$

as well as the discrete spatial spread measures  $M(|h(\cdot, \cdot)|)$  and  $V(|h(\cdot, \cdot)|)$  according to Eqs. (3) and (4) for the 8 “master key filters” extracted by Babaiee et al. [4] and shown in Fig. 1. As can be seen from these results:

- For all the Filters 1–8, the off-diagonal mixed elements in the variance-based spatial spread measures in terms of the covariance matrices  $V(|h(\cdot, \cdot)|)$  are much smaller in magnitude than the diagonal elements. In this respect, the learned receptive field appear to be strongly aligned with the Cartesian coordinate directions.

- For Filters 1–4, one component of the spatial mean value  $M(|h(\cdot, \cdot)|)$  is much smaller in magnitude than the other component, which is rather close to either  $+\frac{1}{2}$  or  $-\frac{1}{2}$ , thus reflecting a substantial deviation from symmetry. In this respect, there is a close structural similarity to the corresponding spatial mean values for the non-centered first-order difference operators in Eqs. (28)–(31). There is, however, also a notable difference in the sense that the mean values do not very closely approximate  $\pm 0.5$ , but rather  $\pm 0.6$  or  $\pm 0.7$ .
- For Filters 5–6 and Filters 7–8, the spatial mean values  $M(|h(\cdot, \cdot)|)$  are rather close to zero, in the sense that their magnitudes are all smaller than 0.1, thus indicating a strong degree of either mirror symmetry or antisymmetry.
- For Filters 5–6, the ratio between the diagonal elements in the variance-based spatial spread measure  $V(|h(\cdot, \cdot)|)$  is very close to either 2 or  $\frac{1}{2}$ , indicating a rather close similarity to corresponding continuous variance-based spatial spread measures for first-order Gaussian derivative ker-

nels in Eqs. (15) and (18). The similarity is, however, not fully perfect, thus reflecting a minor notable difference.

- For Filters 1–4, the ratio between the diagonal elements in the variance-based spatial spread measure  $V(|h(\cdot, \cdot)|)$  is, however, not in any way close to either 2 or  $\frac{1}{2}$ , indicating that a direct transfer of the corresponding expressions (15) and (18) for the continuous first-order Gaussian derivatives does not appear as straightforward.
- For Filter 8, the diagonal components of the variance-based spatial spread measure  $V(|h(\cdot, \cdot)|)$  are very similar, thus with a very good qualitative agreement with Eqs. (7) and (10).
- For Filter 7, the diagonal elements of the variance-based spatial spread measure  $V(|h(\cdot, \cdot)|)$  are somewhat similar, although differing in a clearly notable respect.
- The discrete  $l_1$ -norms of Filters 1–4, which later will be shown to be different instances of the same class of filters, are of a similar order of magnitude. The discrete  $l_1$ -norms of Filters 5–6 are also of a similar magnitude. This property indicates that the data-driven way of determining the 8 “master key filters” can be regarded as reasonably robust between the clustered filters from the same class.
- The discrete  $l_1$ -norms of all the filters  $\{1, 2, 3, 4, 5, 6, 7, 8\}$  are of a roughly similar order of magnitude, while differing between the different classes  $\{1, 2, 3, 4\}$ ,  $\{5, 6\}$ ,  $\{7\}$  and  $\{8\}$ .<sup>5</sup>

Concerning the interpretation of, in particular, the variance-based spatial spread measures for Filters 7-8, it should, however, be noted that the definitions of these spatial spread measures assume a background magnitude value equal to zero, while Filters 7–8 in the set of “master key filters” have background magnitude levels that are significantly different from zero. If we are to use the spatial spread measures for capturing the qualitative shapes of both learned and idealized filters in this way, a potential problem, however, is that the occurrence of spurious non-zero values for the learned filters at image positions, where the their corresponding idealized filters are close to zero or very small, can lead to biases in the spatial spread measures. An explanation for this is that the contribution to the spatial spread measure will be weighted by the distance from the origin for the first-order spatial spread measure and by the distance squared for the second-order spatial spread measure.

Thus, if we would match the parameters of the idealized filters to the shapes of the learned filters by regular spatial

<sup>5</sup> While one could potentially consider normalizing the filters with respect to their  $l_1$ -norms, possibly considering the way that the  $L_1$ -norms of continuous Gaussian derivative kernels differ between different order of spatial differentiation, we will, however, not here normalize the filters with respect to their  $l_1$ -norms, and instead normalize the filters with respect to their responses to monomials, as will be later described in Sect. 3.3.3.

spread measures only, the resulting idealized filters could become spatially too large, thereby leading to substantial bias caused by the spatial spread measures, by substantially overestimating the spatial extent of the filters. Thus, some modification is needed, to be able to apply the variance-based spatial spread measures  $V(|h(\cdot, \cdot)|)$  for estimating the filter parameters in the idealized models to be used for modeling Filters 7–8. We will return to this issue in Sects. 3.4 and 3.5.

### 3.3 Measuring the Accuracy in Relation to Centered or Non-Centered Difference Operators Combined with Spatial Smoothing

A natural consistency requirement, when constructing a centered discrete derivative approximation operator  $\Delta_{x^\alpha y^\beta}$ , that is to approximate the derivative operator  $\partial_{x^\alpha y^\beta}$  of a given order  $(m, n) \in \mathbb{Z}_+^2$  over a 2-D spatial domain, is to require that the result of applying the discrete derivative approximation filter of order  $(m, n)$  applied to a monomial of the same order leads to a response that is spatially constant, with specifically

$$\Delta_{x^\alpha y^\beta}(x^\alpha y^\beta) = \alpha! \beta!, \tag{41}$$

where  $\alpha! = \alpha \cdot (\alpha - 1) \dots 2 \cdot 1$  denotes the factorial of the horizontal derivative order  $\alpha$  and  $\beta! = \beta \cdot (\beta - 1) \dots 2 \cdot 1$  denotes the factorial of the vertical derivative order  $\beta$ . Furthermore, for any integer monomial of order  $(a, b) \in \mathbb{Z}_+^2$ , if either  $a < \alpha$  or  $b < \beta$  (or both), we should require that

$$\Delta_{x^\alpha y^\beta}(x^a y^b) = 0. \tag{42}$$

These consistency requirements are similar to the results of applying corresponding continuous pure derivative operations  $\partial_{x^\alpha y^\beta}$  to corresponding monomials.

#### 3.3.1 Shift-Adjusted Consistency Requirement for Non-Centered Difference Operators Designed to Approximate Spatial Derivative Operators

With regard to modeling the 8 “master key filters” in Fig. 1, we did according to the analysis in connection with Table 1 note that the 8 “master key filters” (i) are not necessarily centered, and (ii) do not decrease as rapidly with increasing distance from the center as the discrete derivative approximation filters considered in Lindeberg [27, 28]. To handle the off-centering of the filters, we do for the purpose of the forthcoming analysis in this paper generalize the criterion in Eq. (41) into the shift-adjusted consistency requirement

$$\Delta_{x^\alpha y^\beta}((x - m_x)^\alpha (y - m_y)^\beta) = \alpha! \beta!. \tag{43}$$

We also generalize the criterion in Eq. (42) into the shift-adjusted consistency requirement

$$\Delta_{x^\alpha y^\beta}((x - m_x)^a (y - m_y)^b) = 0, \tag{44}$$

to hold for any  $(a, b) \in \mathbb{Z}_+^2$ , if either  $a < \alpha$  or  $b < \beta$  (or both), where  $m_x$  and  $m_y$  denote the components in the spatial mean vector of the absolute value of the discrete filter  $h_{x^\alpha y^\beta}(m, n)$ , that implements the computational function of the discrete derivative approximation operator  $\Delta_{x^\alpha y^\beta}$  according to

$$M(|h_{x^\alpha y^\beta}(\cdot, \cdot)|) = \begin{pmatrix} m_x \\ m_y \end{pmatrix}. \tag{45}$$

In the following analysis to be performed, we will specifically evaluate these criteria at the centers of the responses of the discrete derivative approximation filters to different types of polynomials of the form  $(x - m_x)^a (y - m_y)^b$ .

Specifically, these consistency requirements hold for all of the pure non-central or central difference operators,  $\delta_{x+}$ ,  $\delta_{x-}$ ,  $\delta_{y+}$ ,  $\delta_{y-}$ ,  $\delta_x$ ,  $\delta_y$ ,  $\delta_{xx}$ ,  $\delta_{xy}$  and  $\delta_{yy}$ , that we consider in this work.

### 3.3.2 Responses to Monomials for Continuous Gaussian Derivative Operators

Due to the fact that the spatial scale-space representation of any image  $f(x, y)$ , obtained by Gaussian smoothing expressed in terms of a parameterization of the scale parameter in terms of the variance  $s = \sigma^2$  of the Gaussian kernel

$$L(\cdot, \cdot; s) = g(\cdot, \cdot; s) * f(\cdot, \cdot), \tag{46}$$

satisfies the diffusion equation

$$\partial_s L = \frac{1}{2} \nabla^2 L = \frac{1}{2} (L_{xx} + L_{yy}), \tag{47}$$

it follows that the evolution over scale for monomial input  $f_{ab}(x, y) = x^a y^b$  can be described in terms of diffusion polynomials, where we for the lowest orders  $a$  and  $b$  have that

$$f_{00}(x, y; s) = 1, \tag{48}$$

$$f_{10}(x, y; s) = x, \tag{49}$$

$$f_{01}(x, y; s) = y, \tag{50}$$

$$f_{20}(x, y; s) = x^2 + s, \tag{51}$$

$$f_{11}(x, y; s) = x y \tag{52}$$

$$f_{02}(x, y; s) = y^2 + s. \tag{53}$$

**Table 2** Result of computing the responses to the monomials  $\theta_{00} = 1$ ,  $\theta_{10} = x - m_x$  and  $\theta_{01} = y - m_y$  for the 8 “master key filters” shown in Fig. 1, where the spatial offset  $(m_x, m_y)^T = M(|h(\cdot, \cdot)|)$  represents the spatial mean of the absolute value of each filter

Filter	$h_i(\cdot, \cdot) * 1$	$h_i(\cdot, \cdot) * (x - m_x)$	$h_i(\cdot, \cdot) * (y - m_y)$
1	$0 \cdot 10^{-5}$	-0.079	+1.556
2	$0 \cdot 10^{-5}$	-1.380	+0.396
3	$0 \cdot 10^{-5}$	+0.129	-1.091
4	$0 \cdot 10^{-5}$	+0.839	+0.029
5	$0 \cdot 10^{-5}$	+7.750	-0.063
6	$0 \cdot 10^{-5}$	-0.160	+6.137
7	$0 \cdot 10^{-5}$	+0.085	-0.169
8	$0 \cdot 10^{-5}$	-0.072	-0.167

From this summary, we can thus read off that the Gaussian smoothing function will leave the constant input function  $f_{00}(x, y) = 1$  as well as the linear input functions  $f_{10}(x, y) = x$  and  $f_{01}(x, y) = y$  unchanged.

This property will then imply that the responses of the zero-order Gaussian kernel and its first-order derivatives in the  $x$ - and  $y$ -directions are given by

$$g(\cdot, \cdot) * 1 = 1, \tag{54}$$

$$g(\cdot, \cdot) * x = x, \tag{55}$$

$$g(\cdot, \cdot) * y = y, \tag{56}$$

$$g_x(\cdot, \cdot) * 1 = 0, \tag{57}$$

$$g_x(\cdot, \cdot) * x = 1, \tag{58}$$

$$g_x(\cdot, \cdot) * y = 0, \tag{59}$$

$$g_y(\cdot, \cdot) * 1 = 0, \tag{60}$$

$$g_y(\cdot, \cdot) * x = 0, \tag{61}$$

$$g_y(\cdot, \cdot) * y = 1. \tag{62}$$

By superposition, this property will then also apply to any shift-adjusted input polynomial of the form  $f_{ab}(x, y) = (x - m_x)^a (y - m_y)^b$ .

By further differentiating these relationship, it follows that the first-order non-centered Gaussian derivative response in the  $x$ -direction to the monomial input  $f_{10}(x, y) = x - m_x$  as well as the first-order non-centered Gaussian derivative response in the  $y$ -direction to the monomial input  $f_{01}(x, y) = y - m_y$  will be equal to 1, if the centering for the non-centered Gaussian derivative responses is around the point  $(m_x, m_y)^T$ .

**Table 3** Result of computing the responses to the monomials  $\theta_{00} = 1$ ,  $\theta_{10} = x - m_x$  and  $\theta_{01} = y - m_y$  for *normalized versions* of the first 6 “master key filters” shown in Fig. 1, where the spatial offset  $(m_x, m_y)^T = M(|h(\cdot, \cdot)|)$  represents the spatial mean of the absolute value of each filter. For this table, the amplitudes of the original first 6 “master key filters” have been rescaled, to make the response equal to 1 for the monomial of  $\theta_{10} = x - m_x$  or  $\theta_{01} = y - m_y$  that best matches an idealized approximation of the filter as either a derivative in the  $x$ -direction or the  $y$ -direction

Filter	$h_{i,\text{norm}}(\cdot, \cdot) * (x - m_x)$	$h_{i,\text{norm}}(\cdot, \cdot) * (y - m_y)$
1	-0.051	+1.000
2	+1.000	-0.287
3	-0.118	+1.000
4	+1.000	+0.035
5	+1.000	-0.008
6	-0.026	+1.000

### 3.3.3 Responses to Monomials for the 8 “Master Key Filters” Operators

Table 2 shows the result of computing the responses to the monomials  $\theta_{00} = 1$ ,  $\theta_{10} = x - m_x$  and  $\theta_{01} = y - m_y$  for the 8 “master key filters” in Fig. 1, with the spatial offset  $(m_x, m_y)^T$  determined from the spatial mean of the absolute value of each filter according to  $(m_x, m_y)^T = M(|h(\cdot, \cdot)|)$ . As can be seen from these results:

- The responses of the filters to the constant function  $\theta_{00} = 1$  are very close to zero for all the filters, indicating that the filters in this respect can be regarded as DC-balanced.<sup>6</sup>
- For Filters 2, 4 and 5, the largest response is obtained for the monomial  $\theta_{10} = x - m_x$ , indicating that these filters respond as strongest in a way rather closely related to first-order derivatives in the horizontal  $x$ -direction.
- For Filters 1, 3 and 6, the largest response is obtained for the monomial  $\theta_{01} = y - m_y$ , indicating that these filters respond as strongest in a way rather closely related to first-order derivatives in the vertical  $y$ -direction.
- The Filters 1–6, which approximate first-order derivatives in either the horizontal  $x$ -direction or the vertical  $y$ -direction, are, however, not normalized to return

derivative responses to the functions  $\theta_{10} = x - m_x$  and  $\theta_{01} = y - m_y$ , to be equal to one in the respective cases of matching orders between the monomials and the derivative approximations.

- For Filters 7 and 8, which have the qualitatively similar visual appearance as a discrete Gaussian kernel, or a discrete delta function with a subtracted Laplacian-of-the-Gaussian, the DC-balancing of the “master key filters” is, however, not fully consistent with discrete Gaussian models of those filters, since both the continuous integral of the continuous Gaussian kernel and the discrete sum of the filter coefficients in discrete approximations of Gaussian kernels should be equal to 1

Table 3 shows the result of an extension of this analysis, where we have for the first 6 “master key filters” normalized the amplitudes of the receptive fields, to make the response equal to 1 for the monomial of  $\theta_{10} = x - m_x$  or  $\theta_{01} = y - m_y$  that best matches an idealized approximation of the filter as either a derivative in the  $x$ -direction or the  $y$ -direction:

$$h_{1,\text{norm}}(m, n) = \frac{h_1(m, n)}{h_1(\cdot, \cdot) * (y - m_y)|_{(x,y)^T=(0,0)^T}}, \tag{63}$$

$$h_{2,\text{norm}}(m, n) = \frac{h_2(m, n)}{h_2(\cdot, \cdot) * (x - m_x)|_{(x,y)^T=(0,0)^T}}, \tag{64}$$

$$h_{3,\text{norm}}(m, n) = \frac{h_3(m, n)}{h_3(\cdot, \cdot) * (y - m_y)|_{(x,y)^T=(0,0)^T}}, \tag{65}$$

$$h_{4,\text{norm}}(m, n) = \frac{h_4(m, n)}{h_4(\cdot, \cdot) * (x - m_x)|_{(x,y)^T=(0,0)^T}}, \tag{66}$$

$$h_{5,\text{norm}}(m, n) = \frac{h_5(m, n)}{h_5(\cdot, \cdot) * (x - m_x)|_{(x,y)^T=(0,0)^T}}, \tag{67}$$

$$h_{6,\text{norm}}(m, n) = \frac{h_6(m, n)}{h_6(\cdot, \cdot) * (y - m_y)|_{(x,y)^T=(0,0)^T}}. \tag{68}$$

As can be seen from these results:

- The complementary response to a first-order monomial in the direction orthogonal to the preferred direction of the receptive field is comparably rather low (of the order of 5 % or below) for Filters 1, 4, 5 and 6.

In this respect, the responses to first-order monomials are rather well consistent as approximations to first-order derivative operators for Filters 1, 4, 5 and 6.

- For Filter 2, the response to a first-order monomial in the direction orthogonal to the preferred direction of the receptive field is notable (of the order of 12 %) for Filter 3, and very notable (of the order of 30 %) for Filter 2.

In this respect, there are notable deviations from approximations to first-order derivative operators for Filter 3 and specifically for Filter 2.

<sup>6</sup> In the greedy search procedure for extracting the “master key filters”, a normalization of the filter coefficients was used to having their sum equal to zero. In the data format used for storing the “master key filters” to be used as input for the modeling step, the filter coefficients were additionally rounded off to 6 decimals. Thereby, the responses of the filters to the monomial 1 will be quantized in units of  $1 \cdot 10^{-6}$ . In Table 2, we have rounded off all these values to  $0 \cdot 10^{-5}$ , implying consistency with the data-driven way that this analysis has been performed, without any use of such prior information, and the quantization errors and the floating-point errors caused by the finite precision in the storage of the “master key filters” and the computations on them.

To conclude, Filters 1–6 show qualitatively rather good or very good structural similarities to first-order derivative approximation filters along the Cartesian coordinate direction.

The matches to the corresponding fully idealized models in terms of Gaussian derivatives or discrete approximations thereof is, however, far from perfect. A possible source to deviations from the idealized theory could also be because of from variations in the background periphery of the learned filters, which, by a lack of symmetry in those spurious variations, could contribute significantly to the responses to the first-order monomials.

### 3.4 Reducing the Bias in the Estimates of the Spatial Variance of the Gaussian Smoothing-Like Receptive Fields by DC-Compensation

As previously remarked, for Filters 7 and 8, the estimates of the spatial extents of the filters obtained from direct application of the variance-based spatial spread measures will be significantly biased, because of the lack of a determination of the DC-level of these filters to correspond to a background level of about zero. Instead, as we saw from the experimental results in Table 2, the response of these filters to the constant function  $\theta_{00} = 1$  is very close to zero, as opposed to being close to one, as it should be for a continuous Gaussian filter or a discrete approximation thereof.

To address this issue, we will in the following DC-correct Filters 7 and 8 according to

$$h_{7,DC}(m, n; C_7) = h_7(m, n) - C_7, \tag{69}$$

$$h_{8,DC}(m, n; C_8) = h_8(m, n) - C_8, \tag{70}$$

by determining the values of the DC-correction-constants  $\hat{C}_7$  and  $\hat{C}_8$  so as to minimize the determinant of the variance-based spatial spread measure  $V(|h(\cdot, \cdot)|)$  according to

$$\hat{C}_i = \operatorname{argmin}_{C_i} \det V(|h_i(\cdot, \cdot) - C_i|), \tag{71}$$

for  $i \in \{7, 8\}$ , and then further normalize the resulting DC-corrected filters as

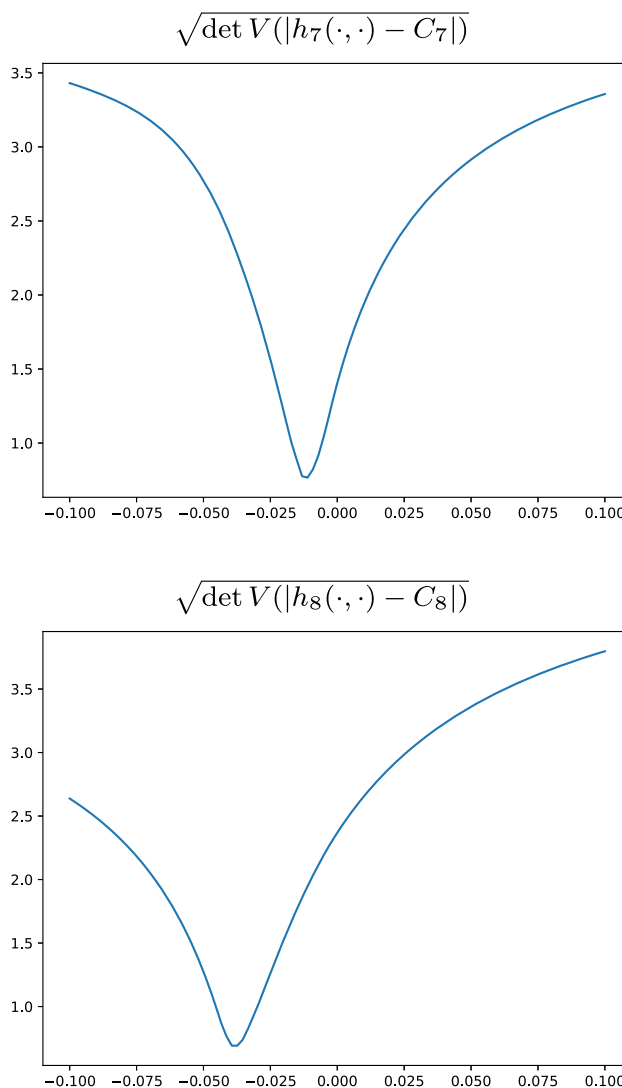
$$h_{7, \text{norm}}(m, n) = \frac{h_{7,DC}(m, n; \hat{C}_7)}{h_{7,DC}(\cdot, \cdot; \hat{C}_7) * 1|_{(x,y)^T=(0,0)^T}}, \tag{72}$$

$$h_{8, \text{norm}}(m, n) = \frac{h_{8,DC}(m, n; \hat{C}_8)}{h_{8,DC}(\cdot, \cdot; \hat{C}_8) * 1|_{(x,y)^T=(0,0)^T}}, \tag{73}$$

so that we obtain the following desirable properties

$$h_{7, \text{norm}}(m, n) * 1 = 1, \tag{74}$$

$$h_{8, \text{norm}}(m, n) * 1 = 1, \tag{75}$$



**Fig. 5** Graphs of the error measure  $\sqrt{\det V(|h_i(\cdot, \cdot) - C_i|)}$  for  $i \in \{7, 8\}$  when determining the DC-compensation constants  $C_7 \approx -0.0118$  and  $C_8 \approx -0.0386$  according to (71), in order to later renormalize the Gaussian-like “master key filters” to unit  $l_1$ -norm, as regular Gaussian kernels obey. (Horizontal axes: parameter value  $C_i \in [-0.1, 0.1]$ . Vertical axes: error measure:  $\sqrt{\det V(|h_i(\cdot, \cdot) - C_i|)}$ )

in analogy to corresponding relationships for continuous Gaussian kernels and appropriate discrete approximations thereof.

Figure 5 show graphs of the error measure

$$\sqrt{\det V(|h_i(\cdot, \cdot) - C_i|)} \tag{76}$$

according to (71) for determining the DC-compensation constants for  $\hat{C}_7$  and  $\hat{C}_8$  in this way. Table 4 shows the result of computing the spatial spread measures for the resulting DC-compensated filters  $h_{7,DC}(m, n)$  and  $h_{8,DC}(m, n)$  according to (69) and (70). As can be seen from the results, the use of DC-compensation substantially reduces the variance-based

**Table 4** Result of computing the discrete  $l_1$ -norm  $\|h(\cdot, \cdot)\|_1$  as well as the discrete spatial mean value  $M(|h(\cdot, \cdot)|)$  according to Eq. (3) and the discrete spatial variance  $V(|h(\cdot, \cdot)|)$  according to Eq. (4) for the DC-compensated “master key filters” 7 and 8 according to (69) and (70)

Filter	$\ h_{i,DC}(\cdot, \cdot)\ _1$	$M( h_{i,DC}(\cdot, \cdot) )$	$V( h_{i,DC}(\cdot, \cdot) )$
7	2.686	$\begin{pmatrix} +0.050 \\ -0.021 \end{pmatrix}$	$\begin{pmatrix} +0.600 & -0.018 \\ -0.018 & +0.962 \end{pmatrix}$
8	1.055	$\begin{pmatrix} +0.019 \\ +0.028 \end{pmatrix}$	$\begin{pmatrix} +0.696 & +0.028 \\ -0.018 & +0.680 \end{pmatrix}$

spatial spread measures for Filters 7 and 8, compared to the variance-based spread measures of the corresponding not DC-compensated filters shown in Table 1.

Table 5 additionally shows the responses to monomials up to order 1 for the resulting DC-compensated and  $l_1$ -normalized filters  $h_{7,norm}(m, n)$  and  $h_{8,norm}(m, n)$  according to (72) and (73). As can be seen from the table, for the larger-size Gaussian-like Filter 8, the responses to the first-order monomials  $x$  and  $y$  are rather low. For the smaller-size Gaussian-like Filter 7, the deviations of those responses from the ideal value 0 are, however, notable.

The top row in Fig. 2 shows a visualization of the resulting normalized versions of the 8 “master key filters”, which we will then approximate by idealized receptive field models in different ways in Sect. 4. Before proceeding to the modeling stage, we will, however, first in the next section introduce a way of defining spatial spread measures, that are less influenced by spurious variabilities in the peripheries of the learned filters.

### 3.5 Reducing the Bias in the Estimates of the Spatial Variances, by Windowed Estimates of the Spatial Spread Measures

As previously explained toward the end of Sect. 3.2, that due to the spurious variations in the background of the learned filters, there may be substantial biases in the mean-based and variance-based spread measures, because of these spurious variations being weighted by either a linear or a quadratic factor in the first- and second-order moments that define the spatial spread measures. This situation is, thus, in contrast to

**Table 5** Result of computing the responses to the monomials  $\theta_{00} = 1$ ,  $\theta_{10} = x - m_x$  and  $\theta_{01} = y - m_y$  for normalized versions of the last 2 “master key filters” shown in Fig. 1, where the spatial offset  $(m_x, m_y)^T = M(|h(\cdot, \cdot)|)$  represents the spatial mean of the absolute

Filter	$h_{i,norm}(\cdot, \cdot) \star 1$	$h_{i,norm}(\cdot, \cdot) \star (x - m_x)$	$h_{i,norm}(\cdot, \cdot) \star (y - m_y)$
7	1.000	-0.098	-0.272
8	1.000	-0.057	-0.117

previous use of such spatial spread measures in Lindeberg [27, 28] for characterizing the properties of idealized mathematical models of receptive fields with regard to variabilities in their spatial extent under variations of the spatial scale parameter.

The underlying reason to this problem is that the idealized filter models, that we are to replace the learned filters with, do all tend to zero further away from the origin, whereas the learned filters contain distinctly non-zero values for image positions where the corresponding idealized filters with similar qualitative shape would assume values very close to zero. Thus, without compensating for this effect, a direct matching of idealized filter shapes to the learned filters based on the regular spatial spread measures in Eqs. (1)–(4) may overestimate the spatial size of the idealized filters that are to replace the learned filters.

The underlying reason for this is that the spurious variations in the background are all transformed to their absolute values, when computing the spatial spread measures. Specifically, regarding the variance-based spatial spread measure, these absolute values are weighted with the squared distance from the origin. Thereby, such spurious variations may contribute in a different way to the spatial spread measures than their effects in the learned filters, where spurious variations of both positive and negative sign could be expected to in some statistical sense counteract each other.

#### 3.5.1 Weighted Spatial Spread Measures

To reduce this potentially substantial source of bias, we will in this section replace the previous uniformly weighted first- and second-order moments by weighted first- and second-order moments. For this purpose, we will use weight functions determined from discrete approximations of Gaussian functions with similar values of the scale parameters, as we would obtain by estimating the scale parameters of idealized filter models from the spatial variances of the corresponding discrete kernels. Thus, we define weighted spatial mean values and spatial variance of continuous or discrete filter kernels  $h(\cdot, \cdot)$  according to:

$$M_{g(\cdot)}(h(\cdot)) = \frac{\int_{p \in \mathbb{R}^2} p h(p) g(p) dp}{\int_{p \in \mathbb{R}^2} h(p) g(p) dp}, \tag{77}$$

value of each filter. For this table, the amplitudes of the original last 2 “master key filters” have been rescaled, to make the response equal to 1 for the monomial of  $\theta_{00} = 1$  equal to 1

$$V_{g(\cdot)}(h(\cdot)) = \frac{\int_{p \in \mathbb{R}^2} p p^T h(p) g(p) dp}{\int_{p \in \mathbb{R}^2} h(p) g(p) dp} - \left( \frac{\int_{p \in \mathbb{R}^2} p h(p) g(p) dp}{\int_{p \in \mathbb{R}^2} h(p) g(p) dp} \right)^2, \tag{78}$$

over a continuous image domain and according to

$$M_{T(\cdot)}(h(\cdot)) = \frac{\sum_{p \in \mathbb{R}^2} p h(p) T(p) dp}{\sum_{p \in \mathbb{R}^2} h(p) T(p) dp}, \tag{79}$$

$$V_{T(\cdot)}(h(\cdot)) = \frac{\sum_{p \in \mathbb{R}^2} p p^T h(p) T(p)}{\sum_{p \in \mathbb{R}^2} h(p) T(p)} - \left( \frac{\sum_{p \in \mathbb{R}^2} p h(p) T(p)}{\sum_{p \in \mathbb{R}^2} h(p) T(p)} \right)^2, \tag{80}$$

over a discrete image domain, with the intention that peripheral values corresponding to spurious structures in the learned filters should have lower influence on the spatial spread measures.

Specifically, if the continuous receptive field  $h(\cdot)$  would be equal to the continuous weighting filter  $g(\cdot)$ , then the weighted spatial mean value and the weighted spatial variance would correspond to the mean value and the standard deviation in  $L_2$ -norm. Similarly, if the discrete receptive field  $h(\cdot)$  would be equal to the discrete weighting filter  $T(\cdot)$ , then the weighted spatial mean value and the weighted spatial variance would correspond to the mean value and the standard deviation in  $l_2$ -norm. For our purpose, we do on the other hand formulate a hybrid approach, where we use the desired shape of the filter model as a prior to suppress the influence of spurious responses in the learned filter spatial positions in the peripheries of the support regions of the learned filters.

### 3.5.2 Determining Idealized Shapes for the Weighting Kernels from the Qualitative Shapes of the “Master Key Filters”

The intention is then to choose either continuous or discrete weighting filters that match the qualitative shapes of the filters to be modeled in the following ways:

- For Filter 8, we will either use anisotropic continuous Cartesian-aligned anisotropic Gaussian kernels

$$g(x, y; \sigma_x, \sigma_y) = g_{1D}(x; \sigma_x) g_{1D}(y; \sigma_y) \tag{81}$$

or discrete Cartesian-aligned anisotropic Gaussian kernels

$$T(m, n; \sigma_x, \sigma_y) = T_{1D}(m; \sigma_x) T_{1D}(n; \sigma_y) \tag{82}$$

having possibly different values of the scale parameters  $\sigma_x$  and  $\sigma_y$  in the  $x$ - and  $y$ -directions, respectively.

- For Filters 2, 4 and 5, we will either use the absolute value of either a first-order horizontal  $x$ -derivative of a continuous Gaussian kernel

$$g_x(x, y; \sigma_x, \sigma_y) = g_{1D,x}(x; \sigma_x) g_{1D}(y; \sigma_y) \tag{83}$$

or the absolute value of a discrete approximation of a horizontal  $x$ -derivative of a Gaussian kernel

$$(\Delta_x T)(m, n; \sigma_x, \sigma_y) = (\Delta_x T_{1D})(m; \sigma_x) T_{1D}(n; \sigma_y) \tag{84}$$

for appropriately matched difference operators  $\Delta_x \in \{\delta_{x-}, \delta_{x+}, \delta_x\}$  in the horizontal  $x$ -direction, with the scale parameters  $\sigma_x$  and  $\sigma_y$  determined from variance-based spatial spread measures of the corresponding filters.

- For Filters 1, 3 and 6, we will either use the absolute value of a first-order vertical  $y$ -derivative of continuous Gaussian kernel

$$g_y(x, y; \sigma_x, \sigma_y) = g_{1D}(x; \sigma_x) g_{1D,y}(y; \sigma_y) \tag{85}$$

or the absolute value of a discrete approximation of a vertical  $y$ -derivative of a Gaussian kernel

$$(\Delta_y T)(m, n; \sigma_x, \sigma_y) = T_{1D}(m; \sigma_x) (\Delta_y T_{1D})(n; \sigma_y) \tag{86}$$

for appropriately matched difference operators  $\Delta_y \in \{\delta_{y+}, \delta_{y-}, \delta_y\}$  in the vertical  $y$ -direction, with the scale parameters  $\sigma_x$  and  $\sigma_y$  determined from variance-based spatial spread measures of the corresponding filters.

- For Filter 7, we will use either a continuous isotropic <sup>7</sup> sharpening filter of the form

$$T_{\text{sharp}}(x, y; \sigma, \gamma) = 1 - \gamma (\nabla^2 g)(x, y; \sigma, \sigma), \tag{87}$$

for an appropriately determined scale parameter  $\sigma$  with an appropriately determined scaling constant  $\gamma$ , where  $\nabla^2 = \partial_{xx} + \partial_{yy}$  denotes the isotropic Laplacian operator, or a discrete isotropic sharpening filter of the form

<sup>7</sup> The reason for using an isotropic Laplacian operator here is mainly to obtain a minimum number of parameters in the model. More generally, one could also conceive replacing the isotropic Gaussian kernel in this model with a separable anisotropic Gaussian kernel, that makes use of different scale parameters in the horizontal and the vertical directions. Possibly, one could also consider extending such a model by making use of different scaling constants  $\gamma_x$  and  $\gamma_y$ , that weigh the contributions from the terms  $\partial_{xx}(g(x; \sigma_x, g(y; \sigma_y))$  and  $\partial_{yy}(g(x; \sigma_x, g(y; \sigma_y))$ , respectively, from the horizontal vs. the vertical directions. The discrete sharpening filter (88) could possibly also be extended to corresponding anisotropic image operations. We do, however, leave the possible investigations of such design options regarding the idealized filter models to future work.

$$T_{\text{sharp}}(x, y; \sigma, \gamma) = 1 - \gamma \nabla_5^2(T_{\text{ID}}(m; \sigma) T_{\text{ID}}(n; \sigma)), \quad (88)$$

where  $\nabla_5^2 = \delta_{xx} + \delta_{yy}$  denotes the 5-point discrete approximation of the Laplacian operator.

### 3.6 Estimating Variance Parameters of Receptive Fields from Weighted Spatial Spread Measures

When using non-uniform weighting kernels for computing the spatial spread measures, the resulting variance-based measures will, however, be affected by the spatial extent of the weighting kernels. In the continuous case, we specifically have the following relations:

$$M_{g(\cdot, \cdot; \sigma_x, \sigma_y)}(g(\cdot, \cdot; \sigma_x, \sigma_y)) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (89)$$

$$V_{g(\cdot, \cdot; \sigma_x, \sigma_y)}(g(\cdot, \cdot; \sigma_x, \sigma_y)) = \begin{pmatrix} \sigma_x^2/2 & 0 \\ 0 & \sigma_y^2/2 \end{pmatrix}, \quad (90)$$

$$M_{|g_x(\cdot, \cdot; \sigma_x, \sigma_y)|}(|g_x(\cdot, \cdot; \sigma_x, \sigma_y)|) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (91)$$

$$V_{|g_x(\cdot, \cdot; \sigma_x, \sigma_y)|}(|g_x(\cdot, \cdot; \sigma_x, \sigma_y)|) = \begin{pmatrix} 3\sigma_x^2/2 & 0 \\ 0 & \sigma_y^2/2 \end{pmatrix}, \quad (92)$$

$$M_{|g_y(\cdot, \cdot; \sigma_x, \sigma_y)|}(|g_y(\cdot, \cdot; \sigma_x, \sigma_y)|) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (93)$$

$$V_{|g_y(\cdot, \cdot; \sigma_x, \sigma_y)|}(|g_y(\cdot, \cdot; \sigma_x, \sigma_y)|) = \begin{pmatrix} \sigma_x^2/2 & 0 \\ 0 & 3\sigma_y^2/2 \end{pmatrix}. \quad (94)$$

Our intention is then to use these expressions for the weighted variance-based spatial spread measures for estimating the scale parameters  $\sigma_x$  and  $\sigma_y$  in idealized models of the learned “master key filters”, based on weighted variance measurements for the learned filters.

Table 6 shows weighted spatial spread measures computed in this way, with the weighting functions for the respective normalized versions of the “master key filters” according to (63)–(68) and (73) determined from the absolute values of the corresponding idealized filter models according to (95)–(102), with the scale parameters of the weighting functions set to  $\sigma_x = \sigma_y = \sigma_0 = 1$  in the spatial weighting kernels for all the “master key filters”.

As can be seen from a comparison with the corresponding non-weighted spatial spread measures in Tables 1 and 4, the use of spatial weighting in the computation of the variance-based spread measures leads to substantially lower estimates of the corresponding scale parameters, and thus to a substantial decrease in the previous bias, caused by spurious variations in the peripheries of the learned filters.

We can also see that the elements of the weighted mean value vectors are, as for the previously computed non-weighted mean values, either around  $\pm 1/2$  or 0, although

**Table 6** Result of computing the discrete *weighted* spatial mean value  $M_{T(\cdot, \cdot)}(|h(\cdot, \cdot)|)$  according to Eq. (79) and the discrete *weighted* spatial variance  $V_{T(\cdot, \cdot)}(|h(\cdot, \cdot)|)$  according to Eq. (80) for the normalized versions of the 8 “master key filters” according to (63)–(68) and (73) when using matching shapes of the spatial weighting filters  $T(\cdot, \cdot)$  to the qualitative shapes of the “master key filters” according to Sect. 3.5.2

Filter	$M_{T(\cdot, \cdot)}( h_i(\cdot, \cdot) )$	$V_{T(\cdot, \cdot)}( h_i(\cdot, \cdot) )$
1	$\begin{pmatrix} -0.005 \\ -0.753 \end{pmatrix}$	$\begin{pmatrix} +0.169 & -0.010 \\ -0.010 & +0.466 \end{pmatrix}$
2	$\begin{pmatrix} +0.665 \\ -0.017 \end{pmatrix}$	$\begin{pmatrix} +0.469 & +0.010 \\ +0.010 & +0.168 \end{pmatrix}$
3	$\begin{pmatrix} +0.011 \\ +0.666 \end{pmatrix}$	$\begin{pmatrix} +0.190 & +0.001 \\ +0.001 & +0.541 \end{pmatrix}$
4	$\begin{pmatrix} -0.536 \\ +0.005 \end{pmatrix}$	$\begin{pmatrix} +0.619 & +0.000 \\ +0.000 & +0.090 \end{pmatrix}$
5	$\begin{pmatrix} -0.011 \\ -0.020 \end{pmatrix}$	$\begin{pmatrix} +1.732 & -0.011 \\ -0.011 & +0.320 \end{pmatrix}$
6	$\begin{pmatrix} -0.006 \\ +0.114 \end{pmatrix}$	$\begin{pmatrix} +0.297 & -0.002 \\ -0.002 & +1.510 \end{pmatrix}$
8	$\begin{pmatrix} -0.002 \\ +0.008 \end{pmatrix}$	$\begin{pmatrix} +0.152 & +0.001 \\ +0.001 & +0.149 \end{pmatrix}$

again with slight offsets from the ideal values by up to 0.1 or 0.2 grid units. In this respect, both the weighted and the unweighted mean values give rise to qualitatively similar results, indicating that the spurious variations in the backgrounds of the “master key filters” seem to have roughly symmetric statistical properties under mirroring through the coordinate axes.

## 4 Modeling the 8 “Master Key Filters” in Terms of Idealized Scale-Space Operations

In the following, we will use the modeling and analysis results from the previous section for modeling the 8 “master key filters” in terms of idealized scale-space operations. Specifically, we will explore the implications of using different types of modeling criteria and analyze how they give rise to different sets of scale parameters in the idealized filter models. Overall, we will either consider separate modeling of the different filters of the following forms:

$$h_{1, \text{ideal}}(m, n; \sigma_{x,1}, \sigma_{y,1}) = (\delta_{y+T})(m, n; \sigma_{x,1}, \sigma_{y,1}), \quad (95)$$

$$h_{2, \text{ideal}}(m, n; \sigma_{x,2}, \sigma_{y,2}) = (\delta_{x-T})(m, n; \sigma_{x,2}, \sigma_{y,2}), \quad (96)$$

$$h_{3, \text{ideal}}(m, n; \sigma_{x,3}, \sigma_{y,3}) = (\delta_{y-T})(m, n; \sigma_{x,3}, \sigma_{y,3}), \quad (97)$$

$$h_{4, \text{ideal}}(m, n; \sigma_{x,4}, \sigma_{y,4}) = (\delta_{x+T})(m, n; \sigma_{x,4}, \sigma_{y,4}), \quad (98)$$

$$h_{5,\text{ideal}}(m, n; \sigma_{x,5}, \sigma_{y,5}) = (\delta_x T)(m, n; \sigma_{x,5}, \sigma_{y,5}), \tag{99}$$

$$h_{6,\text{ideal}}(m, n; \sigma_{x,6}, \sigma_{y,6}) = (\delta_y T)(m, n; \sigma_{x,6}, \sigma_{y,6}), \tag{100}$$

$$h_{7,\text{ideal}}(m, n; \sigma_{x,7}, \sigma_{y,7}) = 1 - \gamma_7 (\nabla_5^2 T)(m, n; \sigma_7, \sigma_7), \tag{101}$$

$$h_{8,\text{ideal}}(m, n; \sigma_{x,8}, \sigma_{y,8}) = T(m, n; \sigma_{x,8}, \sigma_{y,8}), \tag{102}$$

with  $T(m, n; \sigma_x, \sigma_y)$  denoting a 2-D discrete Gaussian kernel with possibly different scale parameters  $\sigma_x$  and  $\sigma_y$  in the two coordinate directions according to (82), and for the special case of Filter 7, the operator  $\nabla_5^2 = \delta_{xx} + \delta_{yy}$  denoting a discrete approximation of the Laplacian operator, and  $\gamma_7 \in \mathbb{R}_+$  denoting a constant to be determined. To keep the dimensionality of the search space down, as well as for theoretical reasons, we have here stated an *a priori* spatially isotropic model for Filter 7 with  $\sigma_{x,7} = \sigma_{y,7} = \sigma_7$ .

The overall form of modeling above is based on the canonical model for spatial receptive fields of the form Lindeberg ([23] Equation 23)

$$T_{\varphi^{m_1} \perp \varphi^{m_2}}(x, y; s, \Sigma) = \partial_{\varphi}^{m_1} \partial_{\perp \varphi}^{m_2} (g(x, y; s \Sigma)), \tag{103}$$

while here (i) restricting the spatial covariance matrix  $\Sigma$  in the affine Gaussian kernel  $g(x, y; s \Sigma)$  to a diagonal matrix according to  $\Sigma_i = \text{diag}(\sigma_{x,i}^2, \sigma_{y,i}^2)$  for each filter  $i$ , and (ii) restricting the directional derivative operators  $\partial_{\varphi}$  and  $\partial_{\perp \varphi}$  to Cartesian partial derivatives  $\partial_x$  and  $\partial_y$ , since the spatial shapes of the learned receptive fields appear to be very much aligned with the Cartesian coordinate directions.

Then, we also make use of the fact that the discrete analog of the Gaussian kernel in (82) and (9) constitutes a canonical way to transfer the scale-space properties of the continuous Gaussian kernel to a discrete domain. The use of complementary centered or non-centered difference operators in (95)–(102) is based on the property that such difference operators applied to a discrete scale-space representation generated by convolution with the discrete analog of the Gaussian kernel preserve discrete scale-space properties [18, 27].

One could, however, also consider using other discrete approximations of Gaussian derivative operators, such as those in Lindeberg [27, 28]

In the following Sects. 4.1–4.6, we will consider different ways of determining the scale parameters  $\sigma_{x,i}$  and  $\sigma_{y,i}$  for  $i \in \{1, 2, 3, 4, 5, 6, 8\}$ , while we will then in Sect. 4.7 consider a different methodology with one more parameter for determining the parameters  $\gamma_7$  and  $\sigma_7$  in an idealized model of Filter 7.

**Note:** In the visualizations of the results for the resulting Methods A–B to be presented, we will, however, not develop specific variance-based methods for Filter 7, since the idealized model of that filter also comprises an amplitude scaling factor  $\gamma_7$  for the contribution from the negative Laplacian-of-the-Gaussian. For this reason, we will for Methods A–B instead interweave visualizations of idealized models of Filter 7 based on the either  $l_2$ -norm-based or  $l_1$ -norm-based Methods D2 and C2, to be described in Sect. 4.7.

### 4.1 Method A: Direct Transfer of Scale Values From Spatial Variances to Idealized Scale-Space Operations with Different Amounts of Smoothing in the Different Coordinate Directions with Use of Continuous Gaussian Derivative Model

A straightforward way of estimating the scale parameters  $\sigma_{x,i}$  and  $\sigma_{y,i}$  in the idealized models (95)–(102) is by making use of the closed-form expressions for the weighted variances of the corresponding continuous Gaussian derivative kernels according to (90), (92) and (94). If we denote the weighted variance matrix for each continuous kernel by

$$V_{|g_{\alpha}(\cdot, \cdot; \sigma_x, \sigma_y)|}(|g_{\alpha}(\cdot, \cdot; \sigma_x, \sigma_y)|) = \begin{pmatrix} v_{xx} & v_{xy} \\ v_{xy} & v_{yy} \end{pmatrix}, \tag{104}$$

then we obtain the following estimates

$$\hat{\sigma}_{x,1} = \sqrt{2 v_{xx,1}}, \quad \hat{\sigma}_{y,1} = \sqrt{2 v_{yy,1}/3}, \tag{105}$$

$$\hat{\sigma}_{x,2} = \sqrt{2 v_{xx,2}/3}, \quad \hat{\sigma}_{y,2} = \sqrt{2 v_{yy,2}}, \tag{106}$$

$$\hat{\sigma}_{x,3} = \sqrt{2 v_{xx,3}}, \quad \hat{\sigma}_{y,3} = \sqrt{2 v_{yy,3}/3}, \tag{107}$$

$$\hat{\sigma}_{x,4} = \sqrt{2 v_{xx,4}/3}, \quad \hat{\sigma}_{y,4} = \sqrt{2 v_{yy,4}}, \tag{108}$$

$$\hat{\sigma}_{x,5} = \sqrt{2 v_{xx,5}/3}, \quad \hat{\sigma}_{y,5} = \sqrt{2 v_{yy,5}}, \tag{109}$$

$$\hat{\sigma}_{x,6} = \sqrt{2 v_{xx,6}}, \quad \hat{\sigma}_{y,6} = \sqrt{2 v_{yy,6}/3}, \tag{110}$$

$$\hat{\sigma}_{x,8} = \sqrt{2 v_{xx,8}}, \quad \hat{\sigma}_{y,8} = \sqrt{2 v_{yy,8}}, \tag{111}$$

with numerical values of these entities shown in the first major column in Table 7 based on the numerical values of the variances in Table 6.

The 2nd row in Fig. 2 shows visualizations of the corresponding idealized receptive fields.

### 4.2 Method B: Matching of Discrete Spatial Variances Between the “Master Key Filters” and the Spatial Variances of Discrete Derivative Approximation Filters, with Individual Scale Parameters for Each Filter

A possible limitation of the above computed estimates of the spatial scale parameters, from weighted variances of the

**Table 7** Summary of the scale estimates obtained for the 8 “master key filters” using the different criteria for fitting the idealized receptive field models to the learned filters: (i) **Method A** based on direct lookup of anisotropic scale estimates from variance-based spatial spread measures, (ii) **Method B** based on anisotropic matching of discrete spatial spread measures between the idealized receptive field models and the learned filters, (iii) **Method C1** based on anisotropic discrete  $l_1$ -

norm minimization, (iv) **Method C2** based on isotropic discrete  $l_1$ -norm minimization, (v) **Method D1** based on anisotropic discrete  $l_2$ -norm minimization and (vi) **Method D2** based on isotropic discrete  $l_1$ -norm minimization. These methods have been explored for all the learned filters except for Filter 7, for which instead the two-parameter modeling scheme in Sect. 4.7, which also determines a scaling constant for the negative Laplacian-of-the-Gaussian filter in the sharpening operation

Filter	Method A		Method B		Method C1		Method C2	Method D1		Method D2
	$\hat{\sigma}_{x,i}$	$\hat{\sigma}_{y,i}$	$\hat{\sigma}_{x,i}$	$\hat{\sigma}_{y,i}$	$\hat{\sigma}_{x,i}$	$\hat{\sigma}_{y,i}$	$\hat{\sigma}_i$	$\hat{\sigma}_{x,i}$	$\hat{\sigma}_{y,i}$	$\hat{\sigma}_i$
1	0.580	0.558	0.644	0.583	0.360	0.510	0.458	0.491	0.722	0.644
2	0.559	0.580	0.586	0.644	0.555	0.453	0.448	0.581	0.519	0.558
3	0.617	0.601	0.690	0.674	0.701	0.655	0.671	0.483	0.503	0.495
4	0.642	0.424	0.756	0.460	0.563	0.384	0.420	0.500	0.000	0.380
5	1.075	0.800	1.107	0.945	1.309	0.875	1.387	1.300	1.004	1.193
6	0.771	1.003	0.900	0.889	0.973	1.171	1.090	0.984	1.074	1.038
7	–	–	–	–	–	–	0.654	–	–	0.675
8	0.552	0.545	0.609	0.601	0.637	0.587	0.611	0.615	0.608	0.612

$$\hat{\sigma}_{x,i} = \operatorname{argmin}_{\sigma_{x,i}} |V_{xx,|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_0, \sigma_0)}(|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})|) - V_{xx,|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_0, \sigma_0)}(|h_{i,\text{norm}}(\cdot, \cdot)|)| \quad (112)$$

$$\hat{\sigma}_{y,i} = \operatorname{argmin}_{\sigma_{y,i}} |V_{yy,|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_0, \sigma_0)}(|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})|) - V_{yy,|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_0, \sigma_0)}(|h_{i,\text{norm}}(\cdot, \cdot)|)| \quad (113)$$

**Fig. 6** Formal specification of the criteria used for determining the scale estimates  $\hat{\sigma}_{x,i}$  and  $\hat{\sigma}_{y,i}$  in the horizontal and vertical directions, by matching the diagonal elements of the discrete variance-based spatial spread measures between an idealized receptive field model  $h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})$  according to (95)–(100) and (102) to the corresponding discrete variance-based spatial spread measures for the normalized versions  $h_{i,\text{norm}}(\cdot, \cdot)$  of the “master key filters” according to (63)–(68) and (73). The parameter  $\sigma_0$  specifies the spatial extent of the discrete weighting kernel used for computing the weighted variance-

based spatial spread measures, and is provided as prior information, chosen as  $\sigma_0 = 1$  for all the filters. (Note that since the idealized receptive field models  $h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})$  are separable with respect to the Cartesian coordinate directions  $x$  and  $y$ , the vertical parameter  $\sigma_{y,1}$  does not influence the estimate of the horizontal diagonal element  $V_{xx,|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_0, \sigma_0)}(|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})|)$ . Similarly, the horizontal parameter  $\sigma_{x,1}$  does not influence the estimate of the vertical diagonal element  $V_{yy,|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_0, \sigma_0)}(|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})|)$ .)

discrete filters, is that the relationship between the weighted variances of the filters is based on a continuous model in terms of Gaussian derivatives,<sup>8</sup> while the filters that we model are genuinely discrete. Especially, since the corresponding scale levels are quite fine, discretization effects in relation to the discrete vs. continuous models may therefore be significant.

To address that issue, we will in this section instead consider a fully discrete approach, where discrete weighted variances of fully discrete models of the idealized receptive fields are matched to discretely computed weighted variances of the learned filters.

Let us denote the operators that extract the diagonal elements in the discrete weighted variance-based spatial spread measure  $V_{T(\cdot)}(h(\cdot))$  in (4) by  $V_{xx,T(\cdot)}(h(\cdot))$  and  $V_{yy,T(\cdot)}(h(\cdot))$ , respectively. In the following, we will determine the scale estimates in idealized models of the “master key filters” according to (95)–(102), based on the criterion that the differences between the diagonal elements of the

variance-based spatial spread measures should be as small as possible, between the idealized receptive field models and the learned filters, as specified in Eqs. (112)–(113) in Fig. 6.

The second major column in Table 7 shows the result of computing horizontal and vertical scale estimates  $\hat{\sigma}_{x,i}$  and  $\hat{\sigma}_{y,i}$  in this way for the 8 different “master key filters”.

The 3rd row in Fig. 2 shows visualizations of the corresponding idealized receptive field models of these filters.

### 4.3 Method C1: $l_1$ -Norm-Based Fitting to Idealized Model for Scale-Space Operations with Different Scale Parameters Along the Horizontal and the Vertical Directions

Although the above spatial spread measure-based modeling operations aim at reflecting the spatial extents of the receptive fields, a conceptual problem arises if there is a mismatch between the shapes of the filters in the idealized receptive fields to the shapes of the learned receptive fields.

In this section, we will address this issue by instead fitting the idealized receptive field models to the learned receptive

<sup>8</sup> According to Eqs. (90), (92) and (94) for continuous Gaussian derivatives, which lead to the estimates in Eqs. (105)–(111).

fields in discrete  $l_1$ -norm. Thus, we will estimate the scale parameters of the receptive fields in the following way:

$$\begin{aligned}
 (\hat{\sigma}_{x,i}, \hat{\sigma}_{y,i}) &= \operatorname{argmin}_{(\hat{\sigma}_{x,i}, \hat{\sigma}_{y,i})} \|\Delta h_{i,\text{norm}}(\cdot, \cdot; \sigma_{x,i}, \sigma_{y,i})\|_1 = \\
 &= \operatorname{argmin}_{(\hat{\sigma}_{x,i}, \hat{\sigma}_{y,i})} \|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,i}, \sigma_{y,i}) - h_{i,\text{norm}}(\cdot, \cdot)\|_1
 \end{aligned}
 \tag{114}$$

with the idealized receptive field models  $h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})$  according to (95)–(102) and the normalized “master key filters”  $h_{i,\text{norm}}(\cdot, \cdot)$  according to (63)–(68) and (73).

The third major column in Table 7 shows the resulting horizontal and vertical scale estimates obtained in this way. The 4th row in Fig. 2 shows visualizations of the corresponding idealized receptive fields.

#### 4.4 Method C2: $l_1$ -Norm-Based Fitting to Idealized Model for Scale-Space Operations with the Same Scale Parameter Along the Horizontal and the Vertical Directions

As a complement to the above analysis, let us also consider the special case when the scale parameters are required to be the same in the horizontal and the vertical directions  $\sigma_{x,i} = \sigma_{y,i} = \sigma_i$ , for which we then determine each scale estimate  $\hat{\sigma}_i$  from the criterion

$$\begin{aligned}
 \hat{\sigma}_i &= \operatorname{argmin}_{\hat{\sigma}_i} \|\Delta h_{i,\text{norm}}(\cdot, \cdot; \sigma_i, \sigma_i)\|_1 = \\
 &= \operatorname{argmin}_{\sigma_i} \|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_i, \sigma_i) - h_{i,\text{norm}}(\cdot, \cdot)\|_1.
 \end{aligned}
 \tag{115}$$

The fourth major column in Table 7 shows the result of computing scale estimates for the idealized receptive field models in this way. The 5th row in Fig. 2 shows visualizations of the resulting receptive fields.

#### 4.5 Method D1: $l_2$ -Norm-Based Fitting to Idealized Model for Scale-Space Operations with Different Scale Parameters Along the Horizontal and the Vertical Directions

In analogy with the above  $l_1$ -norm-based method in Sect. 4.3, for determining separate scale parameters in the horizontal and the vertical directions for idealized receptive field models, we can also use an  $l_2$ -norm-based method from the criterion

$$\begin{aligned}
 (\hat{\sigma}_{x,i}, \hat{\sigma}_{y,i}) &= \operatorname{argmin}_{(\hat{\sigma}_{x,i}, \hat{\sigma}_{y,i})} \|\Delta h_{i,\text{norm}}(\cdot, \cdot; \sigma_{x,i}, \sigma_{y,i})\|_2 = \\
 &= \operatorname{argmin}_{(\hat{\sigma}_{x,i}, \hat{\sigma}_{y,i})} \|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,i}, \sigma_{y,i}) - h_{i,\text{norm}}(\cdot, \cdot)\|_2
 \end{aligned}
 \tag{116}$$

**Table 8** Estimates of the parameters  $\sigma_7$  and  $\gamma_7$  in the sharpening filter model (101) of Filter 7, computed with respect to minimizing either the discrete  $l_1$ -norm or the discrete  $l_2$ -norm of the difference between the idealized receptive field model and the normalized filter (72) according to (118) and (119)

Norm	$\hat{\sigma}_7$	$\hat{\gamma}_7$
$l_1$	0.654	0.522
$l_2$	0.675	0.526

with the idealized receptive field models  $h_{i,\text{ideal}}(\cdot, \cdot; \sigma_{x,1}, \sigma_{y,1})$  again according to (95)–(102) and the normalized “master key filters”  $h_{i,\text{norm}}(\cdot, \cdot)$  according to (63)–(68) and (73).

The fifth major column in Table 7 shows the horizontal and the vertical scale estimates obtained in this way. The 6th row in Fig. 2 shows visualizations of the corresponding idealized receptive fields.

#### 4.6 Method D2: $l_2$ -Norm-Based Fitting to Idealized Model for Scale-Space Operations with the Same Scale Parameter Along the Horizontal and the Vertical Directions

Similarly to the case with  $l_1$ -norm minimization, we can also require the scale parameters in the horizontal and the vertical directions  $\sigma_{x,i} = \sigma_{y,i} = \sigma_i$  to be equal, and determine each scale estimate  $\hat{\sigma}_i$  from the criterion

$$\begin{aligned}
 \hat{\sigma}_i &= \operatorname{argmin}_{\hat{\sigma}_i} \|\Delta h_{i,\text{norm}}(\cdot, \cdot; \sigma_i, \sigma_i)\|_2 = \\
 &= \operatorname{argmin}_{\sigma_i} \|h_{i,\text{ideal}}(\cdot, \cdot; \sigma_i, \sigma_i) - h_{i,\text{norm}}(\cdot, \cdot)\|_2,
 \end{aligned}
 \tag{117}$$

see the sixth major column in Table 7 for resulting the resulting scale estimates for the idealized receptive field models, and the 7th row in Fig. 2 for visualizations of the resulting idealized receptive fields.

#### 4.7 Two-Parameter Modeling of Filter 7 as a Sharpening Filter

For Filter 7, for which the idealized model in (101) comprises a complementary scaling parameter  $\gamma_7$  in addition to the scale parameter  $\sigma_7$ , we determine these two parameters simultaneously, by joint minimization over either the discrete  $l_1$ -norm or the discrete  $l_2$ -norm according to

$$\begin{aligned}
 (\hat{\sigma}_7, \hat{\alpha}_7) &= \operatorname{argmin}_{(\sigma_7, \alpha_7)} \|\Delta h_{7,\text{norm}}(\cdot, \cdot; \sigma_7, \alpha_7)\|_1 = \\
 &= \operatorname{argmin}_{(\sigma_7, \alpha_7)} \|h_{7,\text{ideal}}(\cdot, \cdot; \sigma_7, \alpha_7) - h_{7,\text{norm}}(\cdot, \cdot)\|_1,
 \end{aligned}
 \tag{118}$$

$$(\hat{\sigma}_7, \hat{\alpha}_7) = \operatorname{argmin}_{(\sigma_7, \alpha_7)} \|\Delta h_{7,\text{norm}}(\cdot, \cdot; \sigma_7, \alpha_7)\|_2 =$$

$$= \operatorname{argmin}_{(\sigma_7, \alpha_7)} \|h_{7, \text{ideal}}(\cdot, \cdot; \sigma_7, \alpha_7) - h_{i7 \text{norm}}(\cdot, \cdot)\|_2. \tag{119}$$

These methods are in the graphical illustrations above also referred to as “Method C2” and “Method D2”, respectively.

Table 8 shows numerical values of the resulting parameter estimates  $\hat{\sigma}_7$  and  $\hat{\gamma}_7$  obtained in this way.

Visualizations of the corresponding filter maps are given among the visualizations of the other filter maps for Method C2 and Method D2 in the 5th and 7th rows in Fig. 2, respectively.

## 5 Interpretation of the Analysis and Modeling Results in Terms of Scale-Space Theory for Spatial Receptive Fields

### 5.1 The Spaces Spanned by Linear Combinations of the Receptive Field Responses from the Set of “Master Key Filters”

A notable property of the learned receptive fields is that their orientational preferences are very closely aligned to the directions of Cartesian coordinate directions. In our preliminary investigations in Sect. 3.2, we noted that the unweighted spatial covariance matrices of the absolute values of the learned filter kernels were very close to diagonal (see Table 1). In our follow-up analysis in Sect. 3.5, we also noted that the analogous weighted spatial covariance matrices of the absolute values of the learned filter kernels were also very close to diagonal (see Table 6). This overall property then motivated us to state separable idealized models of the learned receptive fields according to (95)–(102), which we demonstrated in Sect. 4 to provide qualitatively reasonable models of the shapes of the learned filters.

Given that natural image structures usually show a substantial variability with regard to orientations in the image domain, one may then ask how such a variability could be spanned by the deep network, in which the learned filters constitute the primary computational primitives. From the property of the first-order directional derivative operator  $\partial_\varphi$  in the direction  $\varphi$

$$\partial_\varphi = \cos \varphi \partial_x + \sin \varphi \partial_y, \tag{120}$$

it does, however, follow that the computation of derivatives in the horizontal and vertical directions, as performed by the operators  $\partial_x$  and  $\partial_y$ , can by linear combinations span the space of all directions  $\varphi$  in the 2-D image plane. Thus, since the shapes of Filters 1–6 can be qualitatively well described by first-order Gaussian derivative kernels along the Cartesian coordinate directions, we could regard the depthwise combinations of the responses from the learned filters as having the potential ability to span a reasonable approximation of the

space of the first-order directional derivatives in any direction  $\varphi$ , at two separate scales: both (i) at the joint scale of Filters 1–4 and (ii) at the joint scale of Filters 5–6.

Secondly, one may ask why there is a clear preference to first-order spatial derivatives in the learned “master key filters”, and not regarding *e.g.* second-order spatial derivatives, and why the receptive fields of Filters 1–4 are not fully centered, but instead shifted by approximately half a grid spacing along its preferred orientation in the image plane. From properties of the non-centered primitive first-order difference operators  $\delta_{x+}$ ,  $\delta_{x-}$ ,  $\delta_{y+}$  and  $\delta_{y-}$ , we can, however, by linear combinations construct their corresponding centered first-order difference operators according to

$$\delta_x = \frac{\delta_{x+} + \delta_{x-}}{2}, \quad \delta_y = \frac{\delta_{y+} + \delta_{y-}}{2}. \tag{121}$$

We can also construct the corresponding centered second-order difference operators along the coordinate directions as

$$\delta_{xx} = \delta_{x+} - \delta_{x-}, \quad \delta_{yy} = \delta_{y+} - \delta_{y-}. \tag{122}$$

Thus, by linear combinations of the output from Filters 1–4, it should be possible to a certain degree of approximation to span the space of second-order directional derivatives along the Cartesian coordinate directions at the joint scale of Filters 1–4.

If we would like to span the space of second-order directional derivatives in any direction  $\varphi$  according to

$$\partial_{\varphi\varphi} = \cos^2 \varphi \partial_{xx} + 2 \cos \varphi \sin \varphi \partial_{xy} + \sin^2 \varphi \partial_{yy}, \tag{123}$$

we would, however, need to have access to a discrete approximation of the mixed second-order derivative operator  $\partial_{xy}$ , in order to be able to compute a reasonable discrete approximation of the second-order directional derivative in any direction in the image plane.

Given this theoretical motivation, one could thus raise the question about whether the performance of depthwise-separable deep networks with the receptive field shapes replaced by idealized receptive field models could be improved by extending the set of “master key filters” with a filter that constitutes a discrete approximation of the mixed second-order Gaussian derivative operator at the joint scale of Filters 1–4:

$$h_{9, \text{ideal}}(m, n; \sigma_{x,9}, \sigma_{y,9}) = (\delta_{xy}T)(m, n; \sigma_{x,9}, \sigma_{y,9}). \tag{124}$$

Concerning these interpretations, it should finally be emphasized that in depthwise-separable CNNs, no direct linear combinations of the depthwise filters are performed. The depthwise convolutions are applied on distinct input feature maps, and the following pointwise layers linearly combine

the outputs of these convolutions. Hence, the interpretations of the receptive field spanning different subspaces should be interpreted in terms of gross properties at the level of the population of the receptive fields only, and not concerning individual receptive fields.

## 5.2 Implications When Designing Gaussian Derivative Networks

In previous studies of Gaussian derivative networks, the receptive fields in these networks have mainly been formulated as linear combinations of spatial derivatives up to orders 2 or 3, and with the underlying Gaussian derivative responses computed at the same spatial scale level, see Jacobsen et al. [13], Lindeberg [24, 25], Pintea et al. [32], Sangalli et al. [34], Penaud-Polge et al. [30], Gavilima-Pilataxi and Ibarra-Fiallo [9] and Perzanowski and Lindeberg [31]. Given the results from the presented analysis of the set of “master key filters”, one may, however, raise the question of whether the performance of Gaussian derivative networks could be improved by:

- Combining receptive field responses from two scale levels instead of one scale level in each Gaussian derivative layer, as motivated by the qualitatively different scale levels of the groups of filters with indices {1, 2, 3, 4, 8} and {5, 6}, respectively.
- Adding a zero-order term to those networks that have not previously made use of such a term, as motivated by the qualitative shape of Filter 8 in the set of “master key filters”.
- Adding an isotropic sharpening term, as motivated by the qualitative shape of Filter 7 in the set of “master key filters”.

In this context, it should be specifically noted that the use of linear combinations of Gaussian derivative responses in the layers of Gaussian derivative network constitutes a computational operation with large structural similarity to the combination of a set of basic filtering primitives in the depthwise combination of layers in depthwise-separable networks.

## 5.3 Implications When Using Idealized Models of “Master Key Filters” in Depthwise-Separable Deep Networks

As previously discussed in Sect. 5.1, it could be interesting to investigate an additional filter, corresponding to a discrete approximation of the mixed second-order derivative  $\partial_{xy}$  at the joint scale of Filters 1–4, could be advantageous to depthwise-separable networks. The reason for this is that it would make it theoretically possible also to a certain degree of approximation span the space of second-order directional

derivative operators according to (123), and in this way enabling formal capture of a significantly richer set of spatial image structures.

As can be seen from the results in Sect. 4, of estimating the scale values of idealized filter models to approximate the 8 “master key filters” by discrete scale-space operations, the numerical values of the resulting scale estimates can, however, vary notably depending on what criteria are used for computing the scale estimates. For Filter 8, the variability depending on the different criteria is marginal. For Filters 1–6, the variability is, however, significant, because of notable deviations between the idealized receptive field models and the learned filters.

One may hence ask what are the possible predictive properties of these different criteria with regard to the application of replacing the learned filter models by idealized receptive fields in depthwise-separable deep networks. Ultimately, the validity of these idealized approximations is determined by the performance in actual experiments.

A natural next step to consider would also be to use the proposed structural forms of idealized receptive fields according to the presented methods as initial seeds, to then optimize the choice of parameter settings for the idealized filters in a complementary training stage, in a corresponding way as learning of the scale levels has been applied in regular Gaussian derivative networks by Pintea et al. [32], Penaud-Polge et al. [30], Saldanha et al. [33], Yang et al. [38], Basting et al. [6] and Perzanowski and Lindeberg [31].

## 6 Experimental Results

To validate the predictive properties of our different idealized filter modeling approaches, we conducted a series of experiments on the ImageNet dataset.<sup>9</sup> Our experimental design follows a similar methodology as described in Babaiee et al. [5], where we systematically replace the learned depthwise filters in DS-CNN architectures with idealized approximations derived from our theoretical analysis.

### 6.1 Comparative Evaluation of Idealized Filter Modeling Methods

We begin by evaluating the predictive performance of all the six modeling methods (Methods A, B, C1, C2, D1, and D2) described in Sect. 4. For each method, we replaced all the depthwise filters in the ConvNeXt V2 Tiny architecture with linear approximations of the form  $af' + b$ , where  $f'$

<sup>9</sup> We used the standard ImageNet-1K dataset, specifically the ILSVRC 2012 (ImageNet Large Scale Visual Recognition Challenge 2012) version, which consists of 1.28 million training images and 50,000 validation images across 1000 object categories.

**Table 9** ImageNet Top-1 accuracy comparison for different idealized filter modeling methods applied to ConvNeXt V2 Tiny without fine-tuning

Method	A	B	C1	C2	D1	D2
Top-1 (%)	63.958	<b>65.700</b>	62.697	60.972	62.330	63.804

Bold font represents the best result.

represents the closest match among our 8 idealized filters, following the procedure outlined in Babaiee et al. [5]

The linear approximation parameters  $a$  and  $b$  were determined by minimizing the Euclidean distance between the original learned filter  $f_c$  and the transformed idealized filter  $af'_c + b$ , using the closed-form solution presented in Eqs. (1)–(3) in the original work. This approach ensures that each learned filter is optimally approximated by a scaled and shifted version of one of our 8 fundamental filter types.

Table 9 presents the ImageNet Top-1 accuracy results for the ConvNeXt V2 Tiny model, when its depthwise filters are replaced with idealized approximations from each of our six methods, without any subsequent fine-tuning of the rest of the parameters.

The results demonstrate that Method B, based on matching discrete weighted variance-based spatial spread measures between idealized receptive field models and learned filters, achieves the highest predictive accuracy at 65.70%. This finding is particularly significant, as it suggests that the discrete modeling approach, which accounts for the inherent discretization effects in the learned filters, provides superior approximation quality, compared to the methods based on continuous Gaussian derivative models or direct norm minimization.

## 6.2 Validation Through Frozen Filter Training

Based on the comparative results, we selected Method B for further investigation of its practical applicability in depthwise-separable networks. We conducted experiments, where the ConvNeXt V2 Tiny model was initialized with our 8 idealized filters (determined using the estimates of the parameters using Method B in Table 7) and trained from scratch on ImageNet with the depthwise filters remaining frozen throughout the 300-epoch training process.

For determining the filter type assignment at initialization, we analyzed the fully trained ConvNeXt V2 Tiny model and classified each depthwise filter according to its closest match among our 8 idealized filter types using the linear approximation procedure described before. These filter type assignments were then fixed throughout the subsequent training process. All the other network parameters were initialized from the FCMAE (fully convolutional masked autoencoder) pre-trained ConvNeXt V2 Tiny weights, following the same

**Table 10** ImageNet Top-1 accuracy comparison for ConvNeXt V2 Tiny with different filter initialization strategies. The results are averaged over 3 independent runs, reporting mean  $\pm$  standard deviation

Configuration	Top-1 Accuracy (%)
Original ConvNeXt V2 Tiny	82.794 $\pm$ 0.091
Frozen 8 master key filters	82.695 $\pm$ 0.021
Frozen 8 filters (Method B)	82.545 $\pm$ 0.029
With learning of the scale parameters	82.609 $\pm$ 0.055

initialization procedure as described in the original ConvNeXt V2 paper. For comparative purposes, we also trained a standard ConvNeXt V2 Tiny model using the FCMAE pre-trained initialization, following the exact hyperparameters and training procedures from the original code repository. This baseline model achieved 82.79% accuracy and serves as our reference for evaluating the performance impact of our idealized filter constraints.

Table 10 compares the performance of different initialization strategies for the ConvNeXt V2 Tiny architecture.

The results reveal a remarkable consistency across different initialization approaches. The model initialized with our Method B idealized filters achieves 82.54% accuracy, representing only a 0.25% decrease compared to the original architecture. This minimal performance degradation, despite using only 8 distinct filter types across all the depthwise layers, provides strong empirical support for our theoretical analysis and demonstrates the effectiveness of our scale-space-based modeling approach.

Notably, the frozen master key filters from Babaiee et al. [5] achieve identical performance to the fully trainable baseline, while our theoretically derived approximations maintain nearly equivalent results. This suggests that our idealized receptive field models successfully capture the essential computational properties of the learned filters.

## 6.3 Learning of the Scale Parameters

To investigate whether the theoretical constraints imposed by our idealized models limit the representational capacity of the network, we conducted an additional experiment, where the 8 filter shapes from Method B were frozen, but their scale parameters  $\sigma_{x,i}$  and  $\sigma_{y,i}$  were made trainable. This approach maintains the theoretical structure of our scale-space operators, while allowing adaptation to the specific requirements of the layers and the architecture.

Since the modified Bessel functions of integer order, underlying our idealized models of the receptive fields, are not available as full-fledged primitives in PyTorch, as needed for backpropagation of the scale parameters when performing learning of the scale values, we replaced the discrete analog of the Gaussian kernel in the idealized filter models

**Table 11** Statistics of learned scale parameters across all the layers for the 8 filter types in terms of the means and the standard deviations over the sets of the learned filters.  $N$  indicates the number of total filter instances using each type

Filter	$N$	$\sigma_x$ mean	$\sigma_x$ std	$\sigma_y$ mean	$\sigma_y$ std
1	164	0.505	0.189	0.538	0.254
2	132	0.535	0.234	0.538	0.319
3	211	0.523	0.195	0.548	0.222
4	167	0.534	0.242	0.549	0.307
5	154	0.625	0.098	0.677	0.176
6	705	0.657	0.132	0.623	0.085
7	915	0.588	0.274	0.589	0.197
8	4176	0.781	0.498	0.818	0.486

(95)–(102) with the sampled Gaussian kernel, and combined this kernel with corresponding difference operators according to the methodology in Lindeberg [28].

The model with learnable scale parameters achieved 82.61% accuracy, representing a slight improvement of 0.06% over the fully frozen configuration. This result suggests that while our theoretical parameter estimates provide reasonable initial approximations, gradient-based optimization can yield marginal performance gains.

To understand how the network adapts the scale parameters during training, we analyzed the learned scale parameter distributions across all the depthwise layers. Table 11 presents aggregated statistics for the scale parameters  $\sigma_x$  and  $\sigma_y$  for each of the 8 filter types after training completion.

A pattern emerges from this analysis: for most filter types (Filters 1–7), the standard deviation of the learned scale parameters is relatively low, indicating that the network converges to similar scale parameter values across different instances of each filter type. This suggests that these computational primitives—corresponding to various derivative operations and sharpening filters—have fairly constrained optimal parameter ranges.

In contrast, Filter 8, which corresponds to the pure discrete Gaussian smoothing kernel  $T(m, n; \sigma_{x,8}, \sigma_{y,8})$ , exhibits substantially higher standard deviations (0.498 and 0.486) while also being the most frequently used filter type (4176 instances). This high variance indicates that the network benefits from having different spatial smoothing rates for the Gaussian smoothing operation across different filter instances.

Notably, despite these variations in the learned scale parameters—particularly the deviation from our theoretical estimates and the high variance in Filter 8—the overall network accuracy remained essentially unchanged (82.61% vs. 82.54%). This suggests that while the specific scale parameters may vary, the fundamental computational structure captured by our 8 idealized filter types is sufficient to

maintain good performance, indicating that the filter shapes themselves are more critical than their precise parameterization.

## 7 Summary and Discussion

We have presented an in-depth analysis of different ways of modeling the 8 “master key filters” extracted from the ConvNeXt deep learning architecture by Babaiee et al. [4]. For this purpose, we have first computed quantitative measures of the spatial extent of the learned filters in terms of both unweighted and weighted spatial spread measures. This has given support for modeling the receptive fields of these filters with separable filtering operations along the Cartesian coordinate directions, using non-centered difference operators for modeling Filters 1–4 and centered difference operators for modeling Filters 5–6, in combination with spatial smoothing operations. In these spatial smoothing operations, we have allowed for using possibly different values of the scale parameters along the different coordinate directions, and possibly different values of the scale parameters for the individual filters.

The presented modeling and analysis results show that the predictive properties of the scale-space framework, to derive canonical models for spatial receptive fields in terms of Gaussian smoothing and Gaussian derivatives, do qualitatively well generalize from the previously axiomatic formulations regarding the first layer of visual processing, to the qualitative shapes of the receptive fields learned in certain families of depthwise-separable deep networks. Conceptually, these results are also compatible with previous applications of the use Gaussian derivative operations as the computational primitives in deep networks, as developed by Jacobsen et al. [13], Lindeberg [24, 25], Pintea et al. [32], Sangalli et al. [34], Penaud-Polge et al. [30], Smets et al. [35], Gavilima-Pilataxi and Ibarra-Fiallo [9] and Perzanowski and Lindeberg [31].

To explore the influence of different choices for formulating mathematical criteria to model the learned receptive fields by idealized scale-space operations, we have specifically explored four main paths of such methods:

- Method A based on measuring the spatial extent of the learned filters by weighted variance-based spatial spread measures and then predicting the values of the scale parameters based on properties of continuous Gaussian derivatives.
- Method B based on matching discrete weighted variance-based spatial spread measures between the idealized receptive field models and the learned filters.
- Methods C1 and C2 based on minimizing the discrete  $l_1$ -norm between the idealized receptive field models and the learned filters.

- Methods D1 and D2 based on minimizing the discrete  $l_2$ -norm between the idealized receptive field models and the learned filters.

As we have found from the resulting experiments, for some of the learned filters, the scale estimates may vary notably depending on the choice of matching criterion, while for Filter 8 the difference is minor. The main reason for this variability in discrepancy is because of varying degrees of agreement between the idealized models and the learned filters. To judge which of these methods should lead to the best predictive properties with regard to the performance, when inserting corresponding idealized filters into depthwise-separable deep networks, this performance should therefore be evaluated experimentally. Specifically, to be less dependent on the choice of mathematical criterion for model fitting, we have proposed to use the presented theory for formulating parameterized primitives for the idealized receptive field models, and then learning these parameters by deep learning, in a corresponding way as learning of the scale levels is used in Gaussian derivative networks.

The experimental results of replacing the 8 “master key filters” in the ConvNeXt V2 Tiny deep network by idealized discrete scale-space filters obtained in this way show that results almost as good are obtained by using the idealized discrete scale-space filters as when using either the 8 “master key filters” or the filters learned by back propagation in the original deep network architecture. Thereby, the results demonstrate that the filters learned in a reduced version of the modern and state-of-the-art deep network architecture ConvNeXt appear to be very similar to the filters obtained by axiomatically determined scale-space theory. Furthermore, we have shown that the additional increase in performance obtained by learning the scale parameters is marginal for the ConvNeXt V2 Tiny model, thus demonstrating the very good predictive ability of Method B when applied to the filters learned in this deep learning architecture.

Notably, the amount of spatial smoothing is very small in the idealized models of the 8 “master key filters”. This use of very fine scale levels is consistent with the results of learning of the scale levels in deep networks based on linear combinations of Gaussian derivative operators, where the results reported in Appendix A.2.1 in the supplementary material of Perzanowski and Lindeberg [31] also lead to very fine scale levels.

In this modeling work, we have throughout modeled this amount of smoothing with spatial filters based on the discrete analog of the Gaussian kernel [17] combined with small support difference operators, motivated by the theoretical results in Lindeberg [27, 28], that demonstrate that this discretization method is the most suitable one for transferring scale-space properties from a continuous to a discrete image domain, and the experimental results in Perzanowski and Lindeberg

[31], that show that this discretization method overall gives the best performance among 5 different discretization methods, when approximating Gaussian derivative operators for application in Gaussian derivative networks.

Concerning the weighted variance-based spread measures for estimating the spatial extents of the kernels, we used throughout this work fixed values of the scale parameters for the spatial weighting functions, coarsely determined from the spatial extents of the learned filters as observed by visual inspection. A more general approach in this respect could instead be to vary these scale parameters, or to try to adapt these scale parameters in an iterative manner.

Another conceptual simplification, that we have done in the above model fitting schemes, is that we have normalized the learned receptive fields, that are approximated by first-order derivatives in the horizontal  $x$ -direction (or  $y$ -direction), to having the same response to the monomial  $x$  (or  $y$ ) as a corresponding first-order derivative derivative  $g_x(x, y; \sigma)$  (or  $g_y(x, y; \sigma)$ ). Furthermore, we determined offset constants for Filters 7–8 to minimize their variance-based spatial spread measures, and normalized their responses to a constant unit function to be equal to 1, to make the resulting normalized filters more similar to either a pure spatial smoothing operation or a local sharpening operation.

In these ways, we reduced the dimensionality of the following model fitting search, although the resulting model fitting for the  $l_1$ - and  $l_2$ -norm-based modeling schemes may possibly not be as accurate, as the model fitting could possibly be, if extended to higher-dimensional searches over the model parameters. A possible extension to future work could therefore involve performing such higher-dimensional optimizations over the spaces of the parameters of the idealized receptive fields.

Let us finally remark that although the work in this paper has been focused on modeling and analyzing the 8 specific “master key filters” extracted by Babaiee et al. [4], we propose that a similar methodology could be used for modeling other sets of “master key filters” obtained from other deep networks as well as biological visual receptive fields recorded by neurophysiological measurements.

**Author Contributions** T.L. performed the model fitting and developed the theory for that. Z.B. and P.M.K. performed the experiments on deep learning. All authors contributed to the writing and approved the final version of the manuscript.

**Funding** Open access funding provided by Royal Institute of Technology.

**Data Availability** The Python code used for (i) modelling the 8 “master key filters” in terms of idealized discrete scale-space filters as well as (ii) performing the experiments with the learned filters in the ConvNeXt V2 Tiny architecture replaced by the idealized scale-space filters is available at GitHub at <https://github.com/tonylindeberg/masterfiltmodel>.

**Conflict of interest** T.L. is a member of the editorial board of Journal of Mathematical Imaging and Vision.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Abramowitz, M., Stegun, I.A. (eds.): Handbook of Mathematical Functions, 55th edn. Applied Mathematics Series. National Bureau of Standards, US Government Printing Office (1964). (Reprinted by Dover Publications)
2. Babaiee, Z., Kiasari, P., Rus, D., Grosu, R.: We need far fewer unique filters than we thought. In NeurIPS 2024 Workshop on Scientific Methods for Understanding Deep Learning, (2024a). <https://openreview.net/pdf?id=RV8oJTkqkj>
3. Babaiee, Z., Kiasari, P.M., Rus, D., Grosu, R.: Unveiling the unseen: identifiable clusters in trained depthwise convolutional kernels. In International Conference for Learning Representations (ICLR 2024), (2024b). preprint at [arXiv:2401.14469](https://arxiv.org/abs/2401.14469)
4. Babaiee, Z., Kiasari, P., Rus, D., Grosu, R.: The master key filters hypothesis: deep filters are general. In AAAI Conference on Artificial Intelligence, (2025a). preprint at [arXiv:2412.16751](https://arxiv.org/abs/2412.16751)
5. Babaiee, Z., Kiasari, P., Rus, D., Grosu, R.: The quest for universal master key filters in DS-CNNs. Proc. Neural Information Processing Systems (NeurIPS 2025), (2025b). preprint at [arXiv:2509.11711](https://arxiv.org/abs/2509.11711)
6. Basting, M., Brintjes, R.-J., Wiedemer, T., Kümmerer, M., Bethge, M., van Gemert, J.: Scale learning in scale-equivariant convolutional networks. Proc. Computer Vision and Computer Graphics Theory and Applications (VISAPP 2024), 567: 574, 2024
7. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In Proc. Computer Vision and Pattern Recognition (CVPR 2017), pages 1800–1807, (2017)
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, Z., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Housley, N.: An image is worth 16x16 words: transformers for image recognition at scale. In International Conference on Learning Representations (ICLR 2021), (2021)
9. Gavilima-Pilataxi, H., Ibarra-Fiallo, J.: Multi-channel Gaussian derivative neural networks for crowd analysis. In Proc. International Conference on Pattern Recognition Systems (ICPRS 2023), pages 1–7, (2023)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In Proc. Computer Vision and Pattern Recognition (CVPR 2016), pages 770–778, (2016)
11. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: efficient convolutional neural networks for mobile vision applications. [arXiv:1704.04861](https://arxiv.org/abs/1704.04861), (2017)
12. Iijima, T.: Basic theory on normalization of pattern (in case of typical one-dimensional pattern). Bull. Electrotech. Lab. **26**, 368–388 (1962). ((in Japanese))
13. Jacobsen, J.-J., van Gemert, J., Lou, Z., Smeulders, A.W.M.: Structured receptive fields in CNNs. In Proc. Computer Vision and Pattern Recognition (CVPR 2016), pages 2610–2619, (2016)
14. Koenderink, J.J.: The structure of images. Biol. Cybern. **50**(5), 363–370 (1984)
15. Koenderink, J.J., van Doorn, A.J.: Generic neighborhood operators. IEEE Trans. Pattern Anal. Mach. Intell. **14**(6), 597–605 (1992)
16. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
17. Lindeberg, T.: Scale-space for discrete signals. IEEE Trans. Pattern Anal. Mach. Intell. **12**(3), 234–254 (1990)
18. Lindeberg, T.: Scale-Space Theory in Computer Vision. Springer, (1993a)
19. Lindeberg, T.: Discrete derivative approximations with scale-space properties: a basis for low-level feature extraction. J. Math. Imaging Vis. **3**(4), 349–376 (1993)
20. Lindeberg, T.: On the axiomatic foundations of linear scale-space. In J. Sparring, M. Nielsen, L. Florack, and P. Johansen, editors, Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory, pages 75–97, Copenhagen, Denmark, (1996). Springer
21. Lindeberg, T.: Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. J. Math. Imaging Vis. **40**(1), 36–81 (2011)
22. Lindeberg, T.: A computational theory of visual receptive fields. Biol. Cybern. **107**(6), 589–635 (2013)
23. Lindeberg, T.: Normative theory of visual receptive fields. Heliyon **7**(1), e05897 (2021). <https://doi.org/10.1016/j.heliyon.2021.e05897>
24. Lindeberg, T.: Scale-covariant and scale-invariant Gaussian derivative networks. In A. Elmoataz, A., Fadili, J., Quéau, Y., Rabin, J., and Simon, L. (eds.), Proc. Scale Space and Variational Methods in Computer Vision (SSVM 2021), volume 12679 of Springer LNCS, pages 3–14, (2021b)
25. Lindeberg, T.: Scale-covariant and scale-invariant Gaussian derivative networks. J. Math. Imaging Vis. **64**(3), 223–242 (2022)
26. Lindeberg, T.: A time-causal and time-recursive scale-covariant scale-space representation of temporal signals and past time. Biol. Cybern. **117**(1–2), 21–59 (2023)
27. Lindeberg, T.: Discrete approximations of Gaussian smoothing and Gaussian derivatives. J. Math. Imaging Vis. **66**(5), 759–800 (2024)
28. Lindeberg, T.: Approximation properties relative to continuous scale space for hybrid discretizations of Gaussian derivative operators. Front. Signal Process. (2025). <https://doi.org/10.3389/frsip.2024.1447841>
29. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11976–11986, (2022)
30. Penaud-Polge, V., Velasco-Forero, S., Angulo, J.: Fully trainable Gaussian derivative convolutional layer. In International Conference on Image Processing (ICIP 2022), pages 2421–2425, (2022)
31. Perzanowski, A., Lindeberg, T.: Scale generalisation properties of extended scale-covariant and scale-invariant Gaussian derivative networks on image datasets with spatial scaling variations. J. Math. Imaging Vis. **67**(3), 29 (2025)
32. Pinteá, S.L., Tömen, N., Goes, S.F., Loog, M., van Gemert, J.C.: Resolution learning in deep convolutional networks using scale-space theory. IEEE Trans. Image Process. **30**, 8342–8353 (2021)
33. Saldanha, N., Pinteá, S.L., van Gemert, J.C., Tömen, N.: Frequency learning for structured CNN filters with Gaussian fractional derivatives. [arXiv:2111.06660](https://arxiv.org/abs/2111.06660), (2021)

34. Sangalli, M., Blusseau, S., Velasco-Forero, S., Angulo, J.: Scale equivariant U-net. In Proc. British Machine Vision Conference (BMVC 2022), page 763, (2022)
35. Smets, B.M.N., Portegies, J., Bekkers, E.J., Duits, R.: PDE-based group equivariant convolutional neural networks. *J. Math. Imaging Vis.* **65**(1), 209–239 (2023)
36. Weickert, J., Ishikawa, S., Imiya, A.: Linear scale-space has first been proposed in Japan. *J. Math. Imaging Vis.* **10**(3), 237–252 (1999)
37. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. In Computer Vision and Pattern Recognition (CVPR 2023), (2023)
38. Yang, Y., Dasmahapatra, S., Mahmoodi, S.: Scale-equivariant UNet for histopathology image segmentation. *Proc. Machine Learning Research* 194, 130–148, (2022),

**Zahra Babaiee** is a researcher at the Institute of Computer Engineering, TU Wien. Her research interests lie at the intersection of computer vision, bio-inspired neural networks, and embedded systems. Previously, she was a visiting research scholar at MIT CSAIL. Prior to starting her PhD at TU Wien, she completed her Bachelor's degree in Computer Software Engineering at Sharif University of Technology.

**Peyman M. Kiasari** is a doctoral researcher at the Institute of Computer Engineering, TU Wien, where he is a member of the Trustworthy Autonomous Cyber-Physical Systems (TrustACPS) doctoral college. His research explores the generalizability of convolutional filters across diverse domains, with a specific focus on understanding Depthwise-Separable CNNs. He holds a Master's degree from the University of Waterloo and a Bachelor's degree from Sharif University of Technology

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Tony Lindeberg** is a Professor of Computer Science—Computer Vision at KTH Royal Institute of Technology in Stockholm, Sweden. He was born in 1964, received his MSc degree in 1987, his PhD degree in 1991, became docent in 1996, and was appointed professor in 2000. He was a Research Fellow at the Royal Swedish Academy of Sciences between 2000 and 2010. His research interests in computational vision relate to scale-space theory, image features, object recognition, spatio-temporal recognition, video analysis, deep networks and computational modelling of biological vision. He has developed theories and methodologies for continuous and discrete scale-space representation, visual and auditory receptive fields, hierarchical and deep networks, detection of image features and salient image structures, automatic scale selection, scale-covariant and scale-invariant features, affine-covariant and affine-invariant features, affine and Galilean normalization, temporal, spatio-temporal and spectro-temporal scale-space concepts as well as spatial and spatio-temporal image descriptors for image-based recognition. He does also work on computational modelling of hearing and has previously worked on topics in medical image analysis, brain activation and gesture recognition. He is the author of the book *Scale-Space Theory in Computer Vision*