The Master Key Filters Hypothesis: Deep Filters Are General

Zahra Babaiee
TU Vienna & MIT
zbabaiee@mit.edu

Peyman M. Kiasari
TU Vienna
peyman.kiasari@tuwien.ac.at

Daniela Rus
MIT
rus@mit.edu

Radu Grosu
TU Vienna
radu.grosu@uwien.ac.at

Abstract

This paper challenges the prevailing view that convolutional neural network (CNN) filters become increasingly specialized in deeper layers. Motivated by recent observations of clusterable repeating patterns in depthwise separable CNNs (DS-CNNs) trained on ImageNet, we extend this investigation across various domains and datasets. Our analysis of DS-CNNs reveals that deep filters maintain generality, contradicting the expected transition to class-specific features. We demonstrate the generalizability of these filters through transfer learning experiments, showing that frozen filters from models trained on different datasets perform well and can be further improved when sourced from larger, better-performing models. Our findings indicate that spatial features learned by depthwise separable convolutions remain generic across all layers, domains, and architectures. This research provides new insights into the nature of generalization in neural networks, particularly in DS-CNNs, and has significant implications for transfer learning and model design.

Introduction

Understanding neural network generalization mechanisms is crucial in deep learning research [28, 17]. While many studies focus on test accuracies and domain adaptation, we investigate the role of inner structural aspects, specifically in depthwise separable convolutional neural networks (DS-CNNs).

Traditional CNNs develop first-layer filters resembling Gabor functions or color blobs [13], aligning with biological visual systems [15, 2], but deeper layers become more complex. The highly influential work of [26] characterized first-layer filters as "general" and demonstrated deeper layers as "specialized". This led to the widely accepted conclusion that deeper layers become increasingly specialized and significantly influenced neural network theory, as reflected by its more than 10,000 citations in subsequent research.

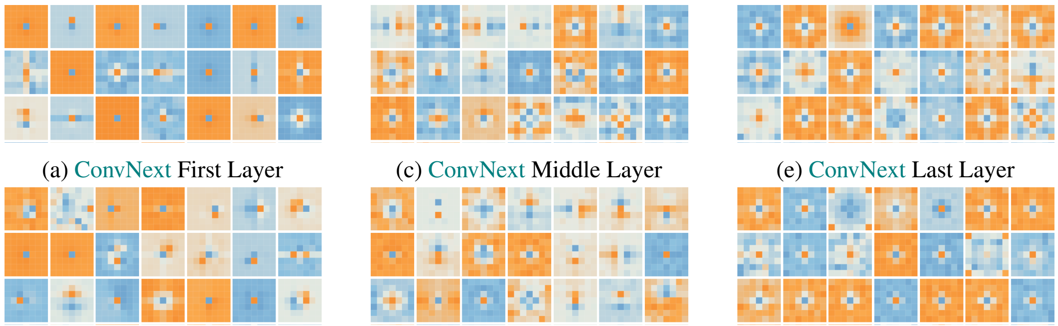
Depthwise separable convolutions decouple spatial feature and channel-wise relationship learning [10, 9], offering a unique perspective on internal spatial information representation. When probing depthwise filters of ImageNet-trained models, repeating patterns emerge across different architectures (Figure 1). Recent research has shown that these filters are clusterable into categories related to Gaussian functions and derivatives [3][4].

Inspired by new observations, we explore the possibility of a generalized function being learned by convolutional networks across *different domains, architectures, and model sizes*. We hypothesize:

The Master Key Filters Hypothesis. *There exist master key filter sets that are general for visual data, and the depthwise filters in DS-CNNs tend to converge to these master key filters, regardless of the specific dataset, task, layer, or architecture.*

To rigorously validate our hypothesis, we conduct experiments across various datasets and domains, recognizing that mere filter clusterability is insufficient proof. Our experiments include:

Workshop on Scientific Methods for Understanding Deep Learning, NeurIPS 2024.



(a) ConvNext First Layer

(c) ConvNext Middle Layer

(e) ConvNext Last Layer

(b) HorNet First Layer

(d) HorNet Middle Layer

(f) HorNet Last Layer

Figure 1: Random depthwise filters sampled from the first, middle, and last layers of ConvNeXt Base and HorNet Small trained on ImageNet. Spatial features in DS-CNNs follow similar patterns regardless of the model architecture.

- *Semantically Divided ImageNet*: Replicating [27]'s experiment, we divide ImageNet into "man-made" and "natural" classes, transferring frozen filters between them.
- *Cross Domain Transfer*: Transferred frozen filters between diverse datasets.
- *Cross Architecture-Domain Transfer*: Transferred frozen filters between models with different architectures and also different training datasets.

Our experiments reveal that depthwise filters in DS-CNNs remain generic across layers and tasks, with spatial features being largely domain-independent. Transfers from diverse datasets improve performance on smaller ones, and these features are architecture-independent, enabling effective cross-domain and cross-architecture filter transfer.

Related Work

Generalization in Deep Learning. Generalization in neural networks has been a key research focus [17, 7]. Various theories explain generalization, but their practical applicability remains debated [28]. Yosinski et al. [26] showed that transferring deeper layers in CNNs degrades performance, suggesting "specilized filter" theory in deep layers.

Depthwise Separable Convolutions. These computationally efficient convolutions have gained popularity [10, 9, 23, 22, 14, 24, 16], enabling lightweight, scalable models. Recent work reveals recurring patterns in depthwise convolutional kernels across DS-CNN models trained on ImageNet [3].

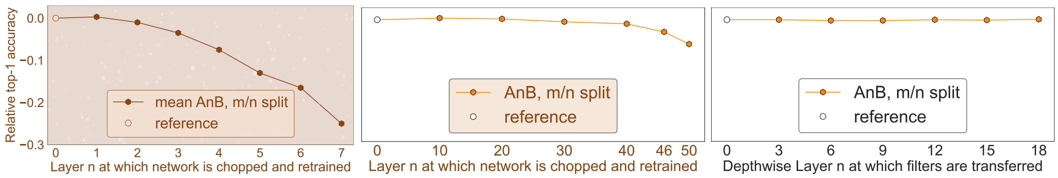
Transfer Learning. Recent work explores initializing large to small models from larger ones [25], analyzes ImageNet transferability [11], and questions pre-training necessity [8].

Generality of spatial features in DS-CNNs

We investigate the universal applicability of spatial features learned by DS-CNNs across diverse domains. Adapting the framework from [26], we define feature generality by their utility in tasks beyond their original scope, examining their effectiveness when transferred from task A to task B. We test our hypothesis through extensive experiments, transferring and freezing depthwise filters from models trained on different datasets.

Revisiting Semantically Divided ImageNet

We replicated the experiment by [26] on convolutional filter transferability across ImageNet subsets (man-made vs. natural objects) using DS-CNNs. We trained Resnet50 on both subsets. We transferred convloutional filters from the first n layers of the man-made model to a new Resnet50 model, froze these layers, and trained on the natural subset. Figure 2b shows the results. We performed the same experiment on ConvNeXt tiny model and transferred depthwise filters from the first n layers of the man-made model to a new ConvNeXt tiny model. Figure 2c shows the results.



(a) Plot Reproduced from [26] shows major performance degradation.

(b) **ResNet50:** Model maintained 92.5% of its original performance.

(c) **ConvNeXt-T:** No performance drop, even at the last layer.

Figure 2: This Figure replicates and extends the study by [26] using depthwise separable convolutions. ImageNet was split into man-made (m) and natural (n) classes. Networks A and B are trained on man-made and natural classes, respectively. The first n layers are transferred from A to B, and this is denoted by AnB. The plots show relative accuracy to base models versus transfer depth. Each point indicates Network B's performance after transferring and freezing filters from A up to layer n, with the remaining layers trained on the natural subset. Notably, depthwise filters exhibit high transferability across all layers, maintaining consistent performance regardless of transfer depth. This suggests a high degree of generality in depthwise convolutional filters, contrasting with traditional CNNs where performance degrades when transferring deeper layers between dissimilar domains.

Contrary to [26]'s findings, our results show no significant performance drop when transferring filters from man-made to natural subset. On Resnet50, we only see visible accuracy drop in the last 10 out of 49 convolutional layers. On ConvNeXt tiny, transferred filters perform comparably to those trained directly on the natural subset, with no substantial performance trend as the number of transferred layers increases. These findings contrast sharply with previous observations in traditional CNNs, where performance degraded as more layers from the dissimilar domain were transferred, especially beyond the first three layers. These findings suggest depthwise convolution filters maintain high generality even in deeper layers, suggesting broad applicability across tasks regardless of layer depth.

Table 1: Accuracy Comparison of Different Filter Transfer Scenarios in ConvNeXt Tiny.

Method	Baseline	Transferred	Shuffle Transferred	Only First 3 layers Transferred
Accuracy	86.9%	86.9%	86.2%	86.9%

To further demonstrate generality, we compared three transfer scenarios against the baseline accuracy: standard transfer (Figure 2c), shuffled filters transfer, and transfer of only the first three layers (repeated to fill all layers). Surprisingly, even extreme scenarios like retaining only the first 3 layers showed no significant accuracy drop as Table 1 shows. These results strongly support the high generality of depthwise convolution filters across various conditions and semantic domains.

Cross Domain Transfer

We examine the cross-domain transferability of depthwise separable convolutional filters using diverse datasets. Our methodology employs ConvNeXt Femto [16] as the base model, training on each dataset, transferring and freezing depthwise filters between dataset pairs, and retraining on target datasets. We also transfer filters from ImageNet-trained models and establish a "selffer" baseline for each dataset, resulting in 42 training configurations (Tables 7 and 2).

Results. Table 2 presents datasets ordered by size, with color-coded performance indicators. Filters from models trained on larger, more diverse datasets consistently enhance performance across domains, suggesting increased data variety fosters more universally applicable filters. No mutual negative impact in cross-dataset filter transfers, further supporting filter generalizability.

These results challenge the notion of strict domain-specificity in depthwise filters and suggest they may capture more general features than previously assumed, with potential implications for transfer learning and model design. (For pointwise layer analysis, see Appendix)

Cross-Architecture Transfer

We investigate filter transferability across architectures and domains using ImageNet and Food 101 as source domains, and Oxford Pets as the target domain. We use ConvNeXt Femto as the base model

Table 2: Accuracy of the ConvNeXt Femto model on the target dataset, with frozen **depthwise filters** transferred from the models trained on source datasets. The diagonal shows the results of the models with "selffer" frozen depthwise filters. The datasets are ordered based on descending training set size. The cell colors red, green, and gray show a decrease, increase, or no change in selffer accuracy, respectively. Arrows indicate relative accuracy (≥ 0.1) compared to the selffer models in each row.

Target \ **Source**	ImageNet	Food	Sketch	Cifar10	STL10	Pets	Flowers
Food	↑87.6%	87.3%	↓86.1%	↓83.6%	↓81.8%	↓78.2%	↓79.1%
Sketch	↑67.1%	66.6%	66.6%	↓65.2%	↓64.1%	↓61.6%	↓61.0%
Cifar10	97.1%	97.0%	97.2%	97.1%	↓96.4%	↓96.0%	↓95.8%
STL10	↑83.2%	↑83.9%	↑84.6%	↑85.2%	82.7%	↓79.3%	↓79.3%
Pets	↑56.0%	↑61.9%	↑63.9%	↑56.8%	↑60.0%	52.4%	↓45.2%
Flowers	↑73.2%	↑71.6%	↑71.4%	↑73.3%	↑71.8%	69.0%	69.1%

and transfer filters from various ConvNeXt and HorNet [21] sizes, which features different structures. For larger source models or those with different channel numbers, we stack and transfer depthwise filters from the beginning. These transferred filters are frozen during training on Oxford Pets.

Results. Table 3 shows that transferring filters from larger ConvNeXt variants and HorNet models trained on ImageNet leads to better accuracy improvements compared to the small femto variant. Notably, HorNet filters perform exceptionally well despite architectural differences.

Table 3: Accuracy of ConvNeXt Femto on the Oxford Pets dataset with transferred filters from different model architectures and sizes trained on ImageNet.

	ConvNeXt			Hornet		
Source Model	Femto	Tiny	Large	Tiny	Small	Large
Accuracy (%)	56.0	67.0	66.5	60.3	63.3	58.8

To further demonstrate filter generality, we transfer filters from HorNet Tiny trained on Food 101 (a distant domain) to ConvNeXt Femto, which is then trained on Oxford Pets. This setup tests transferability across both domains and architectures.

As shown in Table 4, the model achieves 55.5% accuracy, a 3.1% increase over the selfferred baseline. This improvement, despite significant domain and architectural differences, suggests that depthwise convolutions learn general spatial features independent of dataset, domain, and model architecture. These findings indicate that pre-trained filters from various sources can enhance smaller models' performance, even with structural differences, demonstrating the robustness and transferability of learned spatial features in DS-CNNs.

Table 4: Accuracy of ConvNeXt Femto on the Oxford Pets dataset with transferred filters from HorNet Tiny trained on the Food 101 dataset.

Source Model	ConvNext-F on pets	HorNet-T on Foods
Accuracy	52.4%	55.5%

Conclusions

This paper introduces the Master Key Filters Hypothesis. We challenged the prevailing notion that there is a transition from general to specialized filters in CNNs, and filters get increasingly specialized in deeper layers of the network.

References

[1] Keras guide on transfer learning. https://keras.io/guides/transfer_learning/, 2023. Accessed: 2024.

[2] Zahra Babaiee, Ramin Hasani, Mathias Lechner, Daniela Rus, and Radu Grosu. On-off center-surround receptive fields for accurate and robust image classification. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 478–489. PMLR, 18–24 Jul 2021.

[3] Zahra Babaiee, Peyman Kiasari, Daniela Rus, and Radu Grosu. Unveiling the unseen: Identifiable clusters in trained depthwise convolutional kernels. In *The Twelfth International Conference on Learning Representations*, 2024.

[4] Zahra Babaiee, Peyman M. Kiasari, Daniela Rus, and Radu Grosu. Neural echos: Depthwise convolutional filters replicate biological receptive fields. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 8201–8210, 2024.

[5] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. In *European Conference on Computer Vision*, ECCV '14, pages 446–461. Springer, 2014.

[6] Adam Coates, Andrew Y. Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223, 2011.

[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[8] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4918–4927, 2019.

[9] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[10] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861, 2017.

[11] Simon Kornblith, Jonathon Shlens, and Quoc V Le. Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2661–2671, 2019.

[12] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). 2009.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[14] Siyuan Li, Zedong Wang, Zicheng Liu, Cheng Tan, Haitao Lin, Di Wu, Zhiyuan Chen, Jiangbin Zheng, and Stan Z Li. Efficient multi-order gated aggregation network. *arXiv preprint arXiv:2211.03295*, 2022.

[15] Tony Lindeberg. Covariance properties under natural image transformations for the generalised gaussian derivative model for visual receptive fields. *Frontiers in Computational Neuroscience*, 17, June 2023.

[16] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[17] Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester, and Nati Srebro. Exploring generalization in deep learning. *Advances in neural information processing systems*, 30, 2017.

[18] Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *2008 Sixth Indian Conference on Computer Vision, Graphics Image Processing*, pages 722–729, 2008.

[19] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3498–3505, 2012.

[20] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[21] Yongming Rao, Wenliang Zhao, Yansong Tang, Jie Zhou, Ser-Lam Lim, and Jiwen Lu. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[22] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V Le. Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2820–2828, 2019.

[23] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946, 2019.

[24] Asher Trockman and J. Zico Kolter. Patches are all you need? *CoRR*, abs/2201.09792, 2022.

[25] Zhiqiu Xu, Yanjie Chen, Kirill Vishniakov, Yida Yin, Zhiqiang Shen, Trevor Darrell, Lingjie Liu, and Zhuang Liu. Initializing models with larger ones. In *The Twelfth International Conference on Learning Representations*, 2024.

[26] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[27] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. Understanding neural networks through deep visualization. In *International Conference on Machine Learning*, pages 2067–2075, 2015.

[28] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM*, 64(3):107–115, feb 2021.

Appendix / supplemental material

Transfer Method illustration: Depthwise Filter Transfer in DS-CNNs

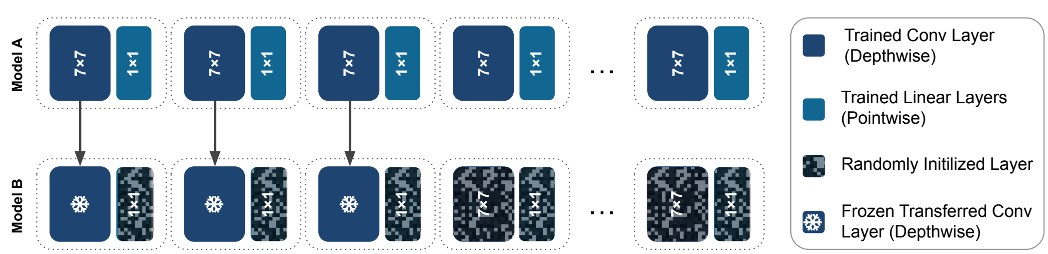


Figure 3: Overview of the experimental setup for depthwise filter transfers. Top: The base model-A is trained on the source dataset-A. Bottom: In the transfer model-B, the first n depthwise convolution layers of the network (in this example, n = 3) are transferred and frozen from the base model-A, the rest of the layers are randomly initialized, and then, they are trained on the target dataset-B. This experiment tests the extent to which the filters on layer n are general or specific.

Are Pointwise Convolutions Specialized?

The results thus far indicate that depthwise filters exhibit significant generality. This raises an intriguing question: If DS-CNNs extract features hierarchically and transition to specialized features, are the pointwise convolutions responsible for this specialization? To address this, we conducted another cross-domain experiment, transferring only the pointwise layers while training the remaining model weights. Table 5 presents the results of these experiments for each pair of datasets.

Surprisingly, transferring pointwise filters consistently decreased accuracy compared to the original model, even in selffer experiments. While improved or maintained accuracy during transfers can suggest filter generality, *the accuracy decreases don't necessarily prove pointwise filter specialization.*

This is particularly evident given that *pointwise filters* transferred from the same dataset also showed significant drops, in contrast to selfferred *depthwise filters*, which generally maintained or improved performance.

Table 5: Accuracy of the ConvNeXt Femto model on the target dataset, with with frozen **pointwise filters** transferred from the models trained on source datasets. The diagonal shows the results of the models with "selffer" frozen depthwise filters. The datasets are ordered based on descending training set size. The cell colors red show a decrease in selffer accuracy compared to the original accuracy, respectively. Arrows indicate relative accuracy compared to the original models in each row.

Target \ **Source**	ImageNet	Food	Sketch	Cifar10	STL10	Pets	Flowers
Food	↓ 62.7%	63.6%	↓64.5%	↓ 62.7%	↓ 60.7%	↓ 59.7%	↓ 59.5%
Sketch	↓45.3%	↓45.0%	47.7%	↓44.9%	↓43.5%	↓43.3%	↓43.1%
Cifar10	↓87.0%	↓86.8%	↓87.5%	87.7%	↓88.0%	↓87.9%	↓86.0%
STL10	↓70.1%	↓73.7%	↓73.7%	↓70.6%	70.5%	↓71.0%	↓67.5%
Pets	↓31.8%	↓41.2%	↓36.6%	↓36.3%	↓39.1%	43.6%	↓33.2%
Flowers	↓52.9%	↓60.0%	↓60.0%	↓52.4%	↓56.0%	↓58.0%	57.0%

The performance degradation observed in these experiments may be attributed to optimization challenges related to splitting networks between co-adapted neurons. This phenomenon, termed "fragile co-adaptation" by [26], suggests that freezing transferred layers may create a loss landscape that hinders optimal filter learning. This difficulty is underscored by the fact that selfferred pointwise filters suffer similarly to those transferred from other domains. Upon examining the filters learned in these experiments, we observed notably noisier patterns, further indicating potential convergence issues.

Is There a Transition from Generic to Class-specific Filters in Deeper Layers?

We investigated the transferability of filters across different layers using ConvNeXt Femto as our base model, with Food 101 and Oxford Pets as source and target datasets. We iteratively transferred layers from models trained on these datasets to new models, as illustrated in Figure 3.

Figure 4 shows the accuracy change for both Food 101-to-Oxford Pets and Oxford Pets-to-Oxford Pets filter transfers, relative to the number of layers transferred. Notably, performance improved as more layers were transferred, stabilizing after 8 layers. This contrasts sharply with traditional convolutions, where performance typically degraded when transferring more layers from a model trained on a distant domain, especially beyond the first three layers.

These findings challenge the conventional understanding of filter specificity in deeper layers and suggest a more nuanced view of filter transferability in modern architectures.

Datasets information

We evaluate transferability across Food 101 [5], Sketch [20], CIFAR-10 [12], Oxford Flowers [18], Oxford Pets [19], and STL-10 [6], representing varied visual domains to challenge filter generalization.

Accuracy of ConvNeXt Femto trained for 300 epochs is available in Table 7

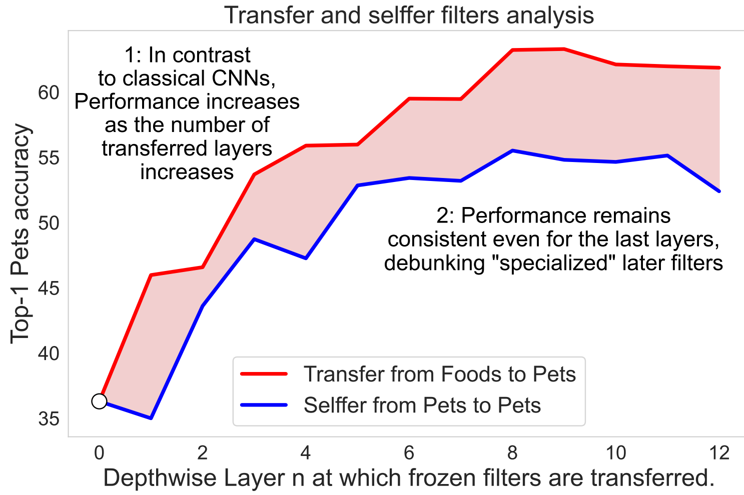


Figure 4: Transferring filters from all layers of a model trained on Foods, improves the performance of the model trained on the distant dataset Pets. Depthwise filters learn general features even in the latest layer, with stark contrast to the task-specialized filters phenomena in traditional CNNs.

Table 6: Dataset information and sample sizes, in training set size descending order.

Dataset	**Classes**	**Train**	**Test**
ImageNet	1000	1.2 m	50 k
Food 101	101	75 k	25 k
Sketch	345	50 k	20 K
CIFAR-10	10	50 k	10 k
STL-10	10	5 k	8 k
Oxford-IIIT Pets	37	4 k	3 k
Oxford 102 Flowers	102	2 k	6 k

Debunking Challenge Submission

What commonly-held position or belief are you challenging?

Provide a short summary of the body of work challenged by your results. Good summaries should outline the state of the literature and be reasonable, e.g. the people working in this area will agree with your overview. You can cite sources beside published work (e.g., blogs, talks, etc).

The highly impactful work of [26], established the widely accepted notion that the filters in deeper layers become increasingly "specialized". This work has been significantly influential, with more than 10K citations. The work characterized the first layer filters of CNNs as "general," and extended its investigation to deeper layers, examining filter generality and specificity through innovative layerwise filter transfer experiments. They empirically demonstrated that when frozen filters from a dissimilar task were transferred, model performance degraded. This led to the widely accepted conclusion of "general" filter layers and "specialized" deeper layers.

This notion is present in many CNN tutorials, especially on transfer learning. It is often suggested that transfer learning can be used to learn new *similar* tasks, for example:[1].

How are your results in tension with this commonly-held position?

Detail how your submission challenges the belief described in (1). You may cite or synthesize results (e.g. figures, derivations, etc) from the main body of your submission and/or the literature.

In this work, we introduced the Master Key Filters Hypothesis for which we provided supporting evidences through a comprehensive set of experiments. We replicated the seminal experiment by[26] on the newer Resnet and ConvNeXt models (See Figure2.) On the ConvNeXt DS-CNN, we observed

Table 7: ConvNeXt Femto model accuracy on the datasets in our benchmark.

Dataset	ImageNet	Food	Sketch	Cifar10	STL10	Pets	Flowers
Accuracy	76.1%	87.6%	66.6%	96.9%	80.4%	36.3%	66.0%

no filter specialization even in the deepest convolutional layers of the network. On the Resnet CNNs, we observed some performance drops when transferring all layers of the network from the dissimialr task. However, the model kept more than 92% of its performance in the worst case.

We further extended our experiments on DS-CNNs to datasets from various domains and models from different architectures. Our findings strongly challenge the notion of transition to specialized filters after the early layers in CNNs. In summary:

1. We demonstrate that depthwise filters in DS-CNNs remain generic across all layers, even for dissimilar tasks.
2. We demonstrate that spatial features learned by DS-CNNs are largely dataset- and domain-independent, and transferring depthwise filters from models trained on datasets with higher sample varieties improves performance on smaller datasets, regardless of domain.
3. Our findings reveal that spatial features in DS-CNNs are independent of model architecture and size, enabling effective cross-architecture filter transfer.
4. We successfully perform cross-domain and cross-architecture filter transfer, further highlighting the generality of DS-CNN spatial features.

How do you expect your submission to affect future work?

Perhaps the new understanding you are proposing calls for new experiments or theory in the area, or maybe it casts doubt on a line of research.

The Master Key Filters Hypothesis advances our understanding of convolutional neural networks and opens new avenues for research, particularly in identifying and leveraging these master filters. Our findings suggest that in case of highly specific data domains, instead of focusing on creating large specialized datasets, we can focus on large, diverse, general datasets. The generality of depthwise filters has implications for transfer learning, enabling performance improvements when transferring filters from larger to smaller datasets, regardless of domain differences. Our results also open new avenues for cross-architecture knowledge transfer. But more important than all, these findings contribute to our understanding of the fundamentals of convolutional neural networks.