The Master Key Filters Hypothesis: Deep Filters Are General

Zahra Babaiee[1*], Peyman M. Kiassari[1*], Daniela Rus[2], Radu Grosu[1]

[1]Technische Universität Wien
[2]Massachusetts Institute of Technology
zahra.babaiee@tuwien.ac.at, peyman.kiasari@tuwien.ac.at, rus@mit.edu, radu.grosu@tuwien.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 filters. 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 datasets. 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 the mechanisms by which neural networks generalize across different tasks and datasets is a pivotal aspect of deep learning research (Zhang et al. 2021; Neyshabur et al. 2017). Generalization, the ability of a model to perform well on unseen data, is often studied by evaluating a model's performance on new, unseen samples or its adaptability to novel domains. While many approaches focus on test accuracies and domain adaptation, in this work, we investigate the role of inner structural aspects of neural networks in generalization, particularly examining the properties of depthwise separable convolutional neural networks (DS-CNNs).

The first layer of traditional convolutional neural networks (CNNs) is well-documented to develop filters resembling Gabor functions or color blobs (Krizhevsky, Sutskever, and Hinton 2012), indicative of their role in capturing basic edge and color information from the visual stimuli. However, as the network progresses into deeper layers, these patterns become more complex and less understood, given the increase in the number of channels and the entangled nature of spatial and channel representations in traditional CNNs.

The highly influential work of (Yosinski et al. 2014) characterized the first layer filters of CNNs as "general," and extended its investigation to deeper layers, examining filter generality and specificity through innovative layerwise feature transfer experiments. They empirically demonstrated that when frozen filters from a dissimilar task were transferred, model performance degradation became progressively more severe as deeper layers were transferred. This led to the widely accepted conclusion that filters in deeper layers become increasingly specialized.

Depthwise separable convolutions are an efficient variant of the standard convolution operation, which decouples the learning of spatial features and channel-wise relationships (Howard et al. 2017, 2019). This separation not only reduces computational complexity but also provides a unique lens through which the internal representation of spatial information can be inspected, even in deep layers of the network.

When probing the depthwise filters of trained models on ImageNet, one observes repeating patterns. Figure 1 shows depthwise filters randomly sampled from the first, middle, and last layer of the trained ConvNeXt (Liu et al. 2022) Base and HorNet (Rao et al. 2022) Small models. The filters have common characteristics between the two different architectures. Recent study has shown that depthwise convolutions across different DS-CNN models trained on the ImageNet dataset are clusterable into distinct categories related to Gaussian functions and derivatives (Babaiee et al. 2024a,b).

Inspired by these observations, our paper seeks to explore the possibility general filter sets being learned by depthwise separable convolutions across *different domains, architectures, and model sizes*.

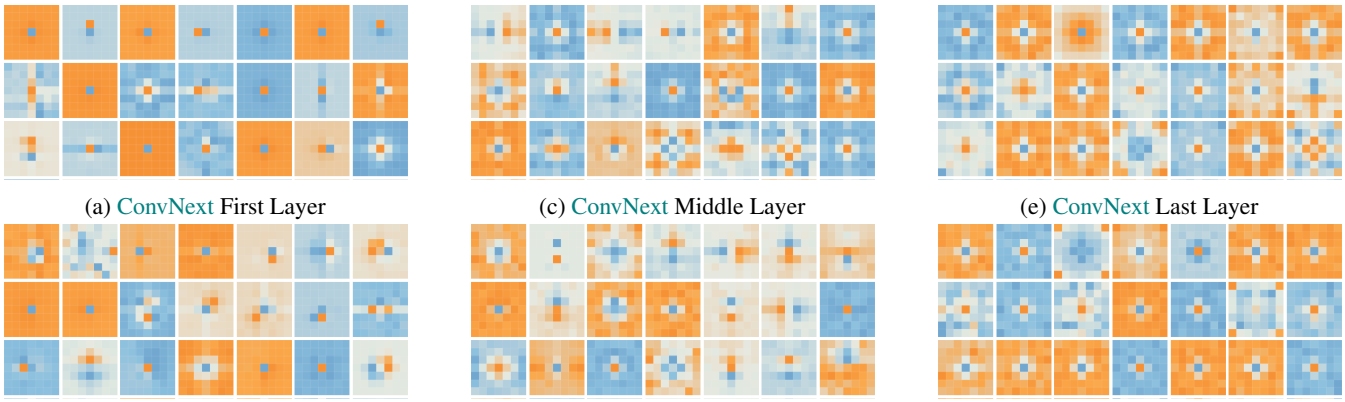
We hypothesize:

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, or architecture.*

To validate this hypothesis, we conduct a comprehensive series of experiments across ImageNet and various other datasets and domains.

1. *Semantically Divided ImageNet:* First, we repeat the well-known experiment from (Yosinski et al. 2015), by dividing ImageNet into "man-made" and "natural" classes that are semantically different from each other. We then transfer and freeze filters from the model trained on man-made

*These authors contributed equally.



(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 and layer.

classes to the model to be trained on natural classes. If our hypothesis is right, unlike (Yosinski et al. 2015), we should see no accuracy drops when transferring deeper layers.

2. *Cross Domain Transfer:* On a set of datasets from various domains, we transfer frozen filters from models trained on each dataset to the other, and investigate the performances.
3. *Cross Architecture Transfer:* We transfer the filters of a model architecture to another distinct model, trained on the same dataset.
4. *Cross Domain and Cross Architecture Transfer:* Finally, we transfer filters from a model with both a different architecture and also trained on a different dataset to another model to be trained on another dataset.

Related Work

Generalization in Deep Learning. Generalization has been a central theme in machine learning research for decades (Neyshabur et al. 2017). The study of generalization seeks to understand how training methodologies, network architectures, and data diversity influence a model's ability to extend beyond its training regime (Goodfellow, Bengio, and Courville 2016). Various theories, such as uniform convergence, margin theory, and algorithmic stability, have been proposed to explain generalization in machine learning. These frameworks often rely on different notions of model complexity, and corresponding generalization bounds quantify the relationship between the amount of data needed and the complexity measure. Despite significant theoretical advancements, the practical value and applicability of these theories remain a subject of ongoing debate in the research community (Zhang et al. 2021). Notably, Yosinski et al. (Yosinski et al. 2014) investigated the transferability of features in deep neural networks by transferring frozen filters from a CNN trained on half of ImageNet to a network to be trained on another half. They showed that transferring deeper than the third layer filters degrades performance, suggesting representation specificity in deep layers.

Depthwise Separable Convolutions. Depthwise separable convolutions have gained popularity over traditional convolutions in recent years due to their computational efficiency and scalabilty (Howard et al. 2017, 2019; Tan and Le 2019; Tan et al. 2019; Li et al. 2022; Trockman and Kolter 2022; Liu et al. 2022). They reduce parameter count and computational complexity by decoupling the spatial and channel computations. These layers have not only facilitated the development of lightweight, scalable models but have also been instrumental in probing the spatial feature extraction capabilities of CNNs. A depthwise-separable convolution is an efficient alternative to standard convolutions in neural networks, splitting the operation into two simpler steps. First, a depthwise convolution applies a separate filter to each input channel independently, capturing spatial patterns within each channel. Mathematically, for an input X with C channels, this performs C separate convolutions: $Y_c = X_c * K_c$, where K_c is the kernel for channel c. Second, a pointwise convolution (1×1 convolution) combines information across channels by applying a $1 \times 1 \times C$ kernel to each spatial location, creating new feature maps: $Z = \sum_{c=1}^{C} Y_c W_c$, where W_c are the weights for each channel. This decomposition significantly reduces the number of parameters and computational cost compared to standard convolutions while maintaining similar expressiveness, making it particularly useful in mobile and edge computing applications. Recent work demonstrates that depthwise convolutional kernels, across various DS-CNN models trained on the ImageNet dataset, exhibit recurring patterns that can be categorized into distinct groups (Babaiee et al. 2024a).

Transfer Learning and Domain Adaptation. Transfer learning focuses on leveraging knowledge from one or more source tasks to improve learning in a related target task. These approaches are particularly valuable in scenarios where labeled data for the target task is scarce or expensive to obtain. (Xu et al. 2024) introduced a method for initializing smaller models by transferring a subset of weights from a pre-trained larger model. Transfer learning does not necessarily improve performance, as transferring knowledge from a dissimilar

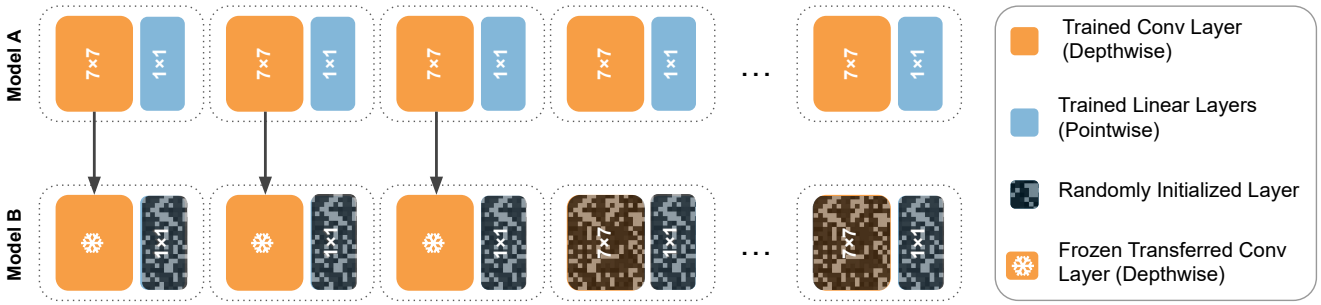


Figure 2: Overview of the experimental setup for depthwise filter transfers. Top: The base model-A is trained on the source dataset-A. Bottom: In the transfer model-B, the first n depthwise convolution layers of the network (in this example, n = 3) are transferred and frozen from the base model-A, the 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.

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

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

domain may not yield positive results (Zhuang et al. 2020). In the realm of computer vision, many studies have investigated the factors that influence the transferability of models trained on ImageNet to other tasks. Kornblith et al. (Kornblith, Shlens, and Le 2019) found that while better ImageNet accuracy generally leads to better transfer performance, this relationship is not always consistent across different architectures and tasks. Additionally, the work by He et al. (He, Girshick, and Dollár 2019) challenges the conventional wisdom of ImageNet pre-training by demonstrating that training models from scratch on target datasets can achieve competitive or even superior performance compared to fine-tuning pre-trained models.

In our work, we transfer *the depthwise filters* and freeze them on a new domain in order to investigate the generality of the features learned across different datasets. Our findings suggest that the spatial features learned by depthwise filters possess a level of generality that allows them to be effectively transferred across different domains.

Generality of Spatial Features in DS-CNNs

Our primary inquiry centers on whether the spatial features learned by DS-CNNs are universally applicable across various datasets and domains. Drawing on the conceptual framework presented in (Yosinski et al. 2014), the generality of learned features can be defined by their utility when applied to tasks beyond their original training purpose. Specifically, examining how effectively these features perform when transferred from their initial training task to a different target task. The feasibility of such transfer relies significantly on the similarity between the source and target tasks.

To rigorously test our hypothesis, we engage in an extensive experimental process where we transfer and freeze the depthwise filters from models trained on different datasets.

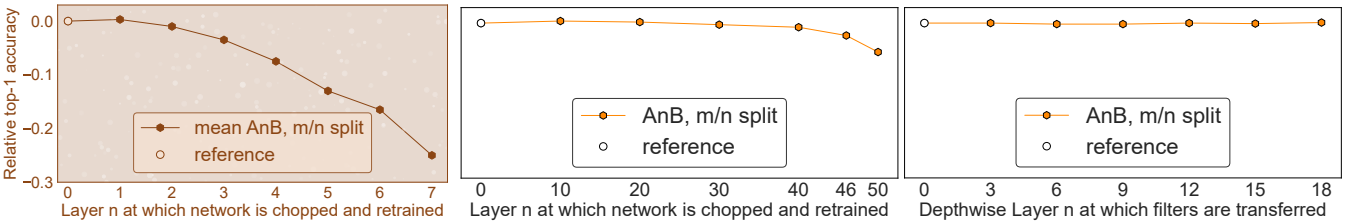
Revisiting Semantically Divided ImageNet

In this section, we replicate the experiment by (Yosinski et al. 2014) on convolutional filter transferability across ImageNet subsets (man-made vs. natural objects) on DS-CNNs. This division creates maximally dissimilar subsets within the ImageNet dataset.

As demonstrated in Figure 2, we transferred the depthwise filters from the first n layers of the model trained on the man-made subset to a new ConvNeXt tiny model. These transferred layers were then frozen, and the model was trained on the natural subset. Figure 3c illustrates the performance results, and Figure 3a shows the exact results re-plotted from (Yosinski et al. 2014). Contrary to (Yosinski et al. 2014), on ConvNeXt, transferred filters perform comparably to those trained directly on the natural subset, with no substantial performance trend as the number of transferred layers increases.

To evaluate the breadth of filter generality, we conducted experiments comparing three distinct transfer scenarios against our baseline accuracy. These scenarios included: (1) standard transfer of all filters (Figure 3c), (2) random shuffling of filters across layers, and (3) a restricted transfer where only the first three layers' filters were used and then repeated throughout the remaining layers. Table 1 shows the results. Surprisingly, even extreme scenarios like retaining only the first 3 layers showed no significant accuracy drop. These results strongly support the high generality of depthwise convolution filters across layers.

These findings raise an important question: Is the enhanced transferability of deeper filters in DS-CNNs, compared to the traditional CNN studied by Yosinski et al., due to the ConvNeXt model's depthwise separable architecture, or do modern training techniques play a crucial role in helping filters learn more generalizable patterns? To investigate this, we conducted the same experiment using ResNet50, a traditional CNN architecture trained with modern methods. The



(a) Plot Reproduced from (Yosinski et al. 2014) shows major performance degradation.

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

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

Figure 3: This Figure replicates and extends the study by (Yosinski et al. 2014) using Resnets and DS-CNNs. 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 2014 experiment where performance degrades when transferring deeper layers between dissimilar domains.

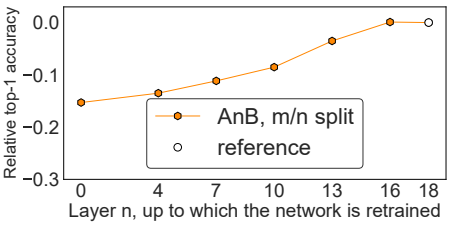


Figure 4: **ResNet18**: Figure 3 experiment, but reversed. Freezing layers from the ***end* to *n*** while training *earlier* layers. Following same reasoning as Yosinski's, these results would suggest later layers are general while initial layers are special.

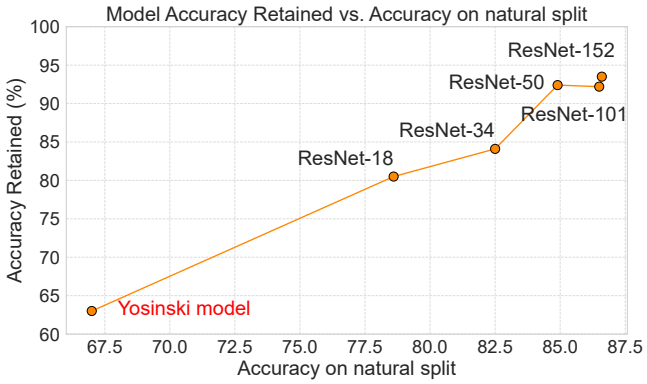


Figure 5: Accuracy retention comparison across ResNet architectures. Deeper networks (ResNet-50/101/152) maintain substantially higher accuracy after transfer compared to shallower networks (ResNet-18/34) and Yosinski's model.

results, shown in Figure 3, reveal that the model stays robust, maintaining 92.5% of it's performance, even when all of it's 49 convolutional layers are transferred.

The contrasting results between Figure 3a and Figure 3b raise another question: Could the performance retention observed in ResNet model be attributed to the residual connections, which were absent in Yosinski's model? Our experimental evidence suggests otherwise. Figure 5 illustrates the accuracy retention across different ResNet architectures following complete layer transfer. We define accuracy retention as the percentage of original accuracy maintained after transfer (e.g., if accuracy decreases from 80% to 70%, the retained accuracy is 70/80=87.5%).

The data reveals a notable pattern: smaller ResNet models exhibit significantly lower accuracy retention, approaching levels similar to those observed in Yosinski's model. This finding has two important insights. First, it challenges the idea that residual connections are strongly responsible for better transfer performance, as smaller ResNets having these connections show poor retention. Second, it also cast new doubts about the "specialized later layers" notion. If layer specialization were an inherent property of CNNs, we would expect to observe increased specialization in models with more layers and, crucially, more pronounced specialization in higher-accuracy models. However, our results demonstrate the opposite trend: deeper ResNet architectures with higher accuracy exhibit more generalized behavior in their later layers.

To further challenge the notion of layer specialization, we conducted the reverse of Yosinski's experiment—freezing layers from the end to n while training earlier layers in Figure 4. Following the same interpretation as Yosinski's framework, these results would paradoxically suggest that later layers are general while initial layers are specialized.

Cross Domain Transfer

This section examines the cross-domain transferability of depthwise separable convolutional filters using a diverse set of datasets varying in size and domain. We aim to assess the generalizability of these filters across disparate datasets.

Datasets. We evaluate the transferability of depthwise filters across six diverse datasets: Food 101 (Bossard, Guillaumin, and Gool 2014)(food images), Sketch (Peng et al. 2019) (hand-drawn object sketches from DomainNet), CIFAR-10 (Krizhevsky 2009) (generic images of vehicles and ani-

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

Table 2: Dataset information and sample sizes, in training set size descending order.

mals), Oxford Flowers (Nilsback and Zisserman 2008) (various flower species), Oxford Pets (Parkhi et al. 2012) (cat and dog breeds), and STL-10(Coates, Ng, and Lee 2011) (mix of animals and objects). The diverse datasets present distinct visual features to test filter transferability. Details of these datasets are summarized in Table 2.

Base Model. Our experiments utilize the ConvNeXt Femto model(Liu et al. 2022). We train the model for 300 epochs in each run, maintaining uniform hyper-parameters across all training instances.

Experimental Setup. Our experimental process is as follows:

1. Train the ConvNeXt Femto model on each of the six datasets separately.
2. For each pair of datasets, transfer all depthwise filters across all layers from the source domain to the network that will be trained on the target dataset and freeze them.
3. Train the model on the target dataset with the transferred and frozen filters.
4. Additionally, for each dataset, transfer the filters of the model trained on ImageNet using the same process.

The models with transferred filters have the advantage of already trained filters. Hence, for a fair comparison of the transferred filters from other domains with the original dataset's accuracy, for each dataset, we also transfer the filters from the model once already trained on it, freeze them, and train the model for another 300 epochs - a baseline termed "selffer" (self-transferred). This process results in 42 different training configurations. The results of these experiments are presented in Tables 3 and 4.

Results. The formatting of Table 4 is particularly illustrative. The table is arranged such that datasets are ordered by their size, with ImageNet, featuring over a million samples, positioned at the forefront, and the Oxford Flowers dataset, which has approximately 2000 samples, at the other end. The diagonal cells denote the "selffer" accuracies. Performance metrics are color-coded to enhance readability and interpretability.

These results reveal key insights, which we will discuss further.

Asymmetric Transfer Effects and Dataset Size: The pattern revealed in Table 4 challenges conventional assumptions about domain-specificity in filter transfer. The predominantly red arrows in the upper triangle and green arrows in the lower triangle, when datasets are sorted by size, indicate that filters from models trained on larger datasets consistently outperform those from smaller ones. This improvement persists regardless of domain similarity between source and target datasets, suggesting that increased data variety leads to more universally applicable filters. Moreover, we never observe mutual negative impact when transferring filters between datasets—a finding that contradicts what might be expected if filters were highly domain-specific. This asymmetric pattern suggests that depthwise filters develop general capabilities that become more robust with increased training variety rather than becoming narrowly specialized to specific domains.

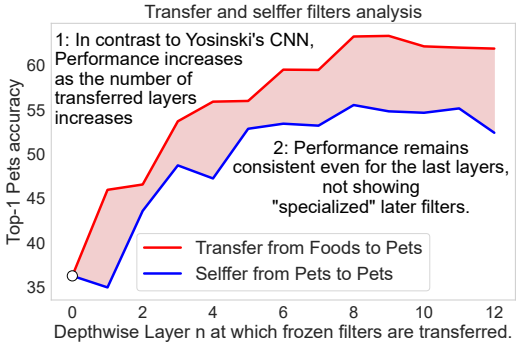


Figure 6: Transferring filters from all layers of a model trained on Foods, improves the performance of the model trained on the distant dataset Pets. Depthwise filters learn general features even in the latest layer, with stark contrast to the task-specialized filters phenomena in Yosinski's CNN.

Is There a Transition from Generic to Class-specific Filters in Deeper Layers? In the experiments in Table 4, we transferred all the filters from all layers to the new models on new domains. To investigate the layer trends, we perform a similar study to the previous section.

We continue to use ConvNeXt Femto as our base model. For the source and target datasets, we use Food 101 and Oxford Pets. Starting from a ConvNeXt Femto model with all layers trained on the Food 101 dataset, we iteratively transfer layers, similar to the procedure shown in Figure 2.

We also repeat the same process by selfferring filters from a ConvNeXt Femto model trained on the Oxford Pets dataset.

Figure 6 shows the accuracy change for both the Food 101-to-Oxford Pets and Oxford Pets-to-Oxford Pets transfer of filters, with respect to the layer number where the source model is chopped. Remarkably, the results demonstrate that the more layers we transfer the filters from, the better the performance becomes. After transferring filters from 8 layers, the performance does not change considerably. These findings too, stand in stark contrast to those previously observed in (Yosinski et al. 2014), where the performance degraded as more layers of the model trained on the far domain were transferred, especially after the first three layers.

What About Pointwise Convolutions, Are They Specialized? The results thus far indicate that depthwise filters exhibit significant generality. This raises an intriguing question:

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

Table 3: ConvNeXt Femto model accuracy on the datasets in our benchmark.

Target \ **Source**	ImageNet	Food	Sketch	Cifar10	STL10	Pets	Flowers
Food	+0.3%	87.3%	-1.2%	-3.7%	-5.5%	-9.1%	-8.2%
Sketch	+0.5%	0.0%	66.6%	-1.4%	-2.5%	-5.0%	-5.6%
Cifar10	0.0%	-0.1%	+0.1%	97.1%	-0.7%	-1.1%	-1.3%
STL10	+0.5%	+1.2%	+1.9%	+2.5%	82.7%	-3.4%	-3.4%
Pets	+3.6%	+9.5%	+11.5%	+4.4%	+7.6%	52.4%	-7.2%
Flowers	+4.1%	+2.5%	+2.3%	+4.2%	+2.7%	-0.1%	69.1%

Table 4: 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 compared to the original accuracy, respectively. Arrows indicate relative accuracy (≥ 0.1) compared to the selffer models in each row.

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.

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 (Yosinski et al. 2014), 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.

Cross-Architecture Transfer

To further investigate the transferability of depthwise filters, we extend our experiments to include different model sizes and architectures, while using the ImageNet dataset as the source domain.

Experimental Setup. For these experiments, we maintain our base model as ConvNeXt Femto and use the Oxford Pets dataset as the target domain. As source models, we use different sizes from the ConvNeXt family (femto, tiny, and large) and introduce another architecture family, HorNet (Rao et al. 2022). HorNet has substantially different blocks compared to ConvNeXt, with recursive gated convolutions. By including HorNet, we can evaluate the transferability of filters from a diverse set of model sizes and architectures.

When a source model is larger than the ConvNeXt Femto or has a different number of channels in its layers, we stack all the depthwise filters in the model and transfer them from the beginning of the stack to the ConvNeXt Femto. These transferred filters are then frozen and used to train the model on the Oxford Pets dataset.

Results. The results of these experiments are presented in Table 6. Transferring filters from larger variants of ConvNeXt trained on ImageNet leads to better accuracy improvements compared to the smaller variant. Interestingly, the transferred filters from HorNet perform exceptionally well, despite being from a completely different model architecture and size.

These results demonstrate that depthwise filters learned by DS-CNNs are highly transferable, with successful transfer even from various architectures despite structural differences. The consistent performance improvements suggest that these spatial features generalize well across datasets, domains, and model architectures.

Cross-Domain and Cross-Architecture Transfer

To further demonstrate the generality of the depthwise filters, we extend our experiments by considering both different domains and different architectures simultaneously. While the ImageNet dataset is large and may already contain features

Target \ Source	ImageNet	Food	Sketch	Cifar10	STL10	Pets	Flowers
Food	−0.9%	63.6%	−1.1%	−0.9%	−2.9%	−3.9%	−4.1%
Sketch	−2.4%	−2.7%	47.7%	−2.8%	−4.2%	−4.4%	−4.6%
Cifar10	−0.7%	−0.9%	−0.2%	87.7%	−0.3%	−0.2%	−1.7%
STL10	−0.4%	−3.2%	−3.2%	−0.1%	70.5%	−0.5%	−3.0%
Pets	−11.8%	−2.4%	−7.0%	−7.3%	−4.5%	43.6%	−10.4%
Flowers	−4.1%	−3.0%	−3.0%	−4.6%	−1.0%	−1.0%	57.0%

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 pointwise 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.

	ConvNeXt			Hornet		
Source Model	Femto	Tiny	Large	Tiny	Small	Large
Accuracy (%)	56.0	+11.0	+10.5	+4.3	+7.3	+2.8

Table 6: Accuracy of ConvNeXt Femto on the Oxford Pets dataset with transferred filters from different model architectures and sizes trained on ImageNet.

useful for classifying pets, we aim to investigate the transferability of filters from a more distant domain. For this purpose, we choose the Food 101 dataset as the source domain, which consists of closeup photos of food on plates or table settings. In contrast, the Oxford Pets dataset, which serves as the target domain, contains images of cat and dog breeds in various settings, such as indoors or outdoors on grass. By selecting these two datasets, we can assess the effectiveness of filter transfer between domains that have minimal common features.

Experimental Setup. We first train the HorNet Tiny model on the Food 101 dataset. We then transfer the depthwise convolutional filters from the trained HorNet Tiny model to the ConvNeXt Femto model, which is subsequently trained on the Oxford Pets dataset with frozen filters. In this scenario, both the dataset domain and the model architecture of the source and target models are different, providing a rigorous test for the generality and transferability of the learned filters.

Results. The results of this experiment are presented in Table 7. Remarkably, the ConvNeXt Femto model trained on the Oxford Pets dataset with transferred filters from the HorNet Tiny model trained on the Food 101 dataset achieves an accuracy of 55.5%, with a 3.1% increase compared to the selfferred baseline. This result is particularly impressive considering the significant differences between the source and target domains, as well as the distinct model architectures.

The results suggest that there are general spatial filter sets learned by depthwise convolutions, regardless of the architecture and dataset domain.

Source Model	ConvNext-F on pets	HorNet-T on Foods
Accuracy	52.4%	+3.1%

Table 7: Accuracy of ConvNeXt Femto on the Oxford Pets dataset with transferred filters from HorNet Tiny trained on the Food 101 dataset.

Conclusions and Discussion

This paper introduces the Master Key Filters Hypothesis, that there exist master key filter sets that are general, and the depthwise filters tend to converge to them. We provide evidence that DS-CNNs learn depthwise convolutional filters that remain general across diverse datasets, domains, and model architectures.

Our experiments, spanning semantically divided ImageNet, cross-domain, and cross-architecture transfers, reveal that these filters maintain generality even in deeper layers, challenging prevailing notions that there is a transition from general to specialized filters, and filters get increasingly specialized in deeper layers of the network.

When transferring the pointwise layers, we observed convergence issues across all datasets, even in selffer models, suggesting optimization difficulties. This may be attributed to the higher parameter count in pointwise layers, potential sparsity effects, and restricted permutation symmetries compared to depthwise layers. These findings align with prior work on fragile co-adaptation in neural networks (Yosinski et al. 2014), where freezing certain layers can create challenging loss landscapes for training the remaining parameters. Future work could further investigate the specific mechanisms behind these optimization challenges.

The generality of depthwise filters has significant 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.

Acknowledgments

Zahra Babaiee and Radu Grosu are supported by the Austrian Science Fund (FWF) project MATTO-GBM I 6605. Peyman M. Kiasari is supported by the TU Wien TrustACPS PhD School program that is supported by TTTech Auto and B&C Privatstiftung.

References

Babaiee, Z.; Kiasari, P.; Rus, D.; and Grosu, R. 2024a. Unveiling the Unseen: Identifiable Clusters in Trained Depthwise Convolutional Kernels. In *The Twelfth International Conference on Learning Representations*.

Babaiee, Z.; Kiasari, P. M.; Rus, D.; and Grosu, R. 2024b. Neural Echos: Depthwise Convolutional Filters Replicate Biological Receptive Fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 8216–8225.

Bossard, L.; Guillaumin, M.; and Gool, L. V. 2014. Food-101 – Mining Discriminative Components with Random Forests. In *European Conference on Computer Vision*, ECCV '14, 446–461. Springer.

Coates, A.; Ng, A. Y.; and Lee, H. 2011. An analysis of single-layer networks in unsupervised feature learning. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 215–223.

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

He, K.; Girshick, R.; and Dollár, P. 2019. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4918–4927.

Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; Le, Q. V.; and Adam, H. 2019. Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; and Adam, H. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR*, abs/1704.04861.

Kornblith, S.; Shlens, J.; and Le, Q. V. 2019. Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2661–2671.

Krizhevsky, A. 2009. Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Li, S.; Wang, Z.; Liu, Z.; Tan, C.; Lin, H.; Wu, D.; Chen, Z.; Zheng, J.; and Li, S. Z. 2022. Efficient multi-order gated aggregation network. *arXiv preprint arXiv:2211.03295*.

Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; and Xie, S. 2022. A ConvNet for the 2020s. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Neyshabur, B.; Bhojanapalli, S.; McAllester, D.; and Srebro, N. 2017. Exploring generalization in deep learning. *Advances in neural information processing systems*, 30.

Nilsback, M.-E.; and Zisserman, A. 2008. Automated Flower Classification over a Large Number of Classes. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722–729.

Parkhi, O. M.; Vedaldi, A.; Zisserman, A.; and Jawahar, C. V. 2012. Cats and Dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3498–3505.

Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019. Moment Matching for Multi-Source Domain Adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Rao, Y.; Zhao, W.; Tang, Y.; Zhou, J.; Lim, S.-L.; and Lu, J. 2022. HorNet: Efficient High-Order Spatial Interactions with Recursive Gated Convolutions. *Advances in Neural Information Processing Systems (NeurIPS)*.

Tan, M.; Chen, B.; Pang, R.; Vasudevan, V.; Sandler, M.; Howard, A.; and Le, Q. V. 2019. Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2820–2828.

Tan, M.; and Le, Q. V. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *CoRR*, abs/1905.11946.

Trockman, A.; and Kolter, J. Z. 2022. Patches Are All You Need? *CoRR*, abs/2201.09792.

Xu, Z.; Chen, Y.; Vishniakov, K.; Yin, Y.; Shen, Z.; Darrell, T.; Liu, L.; and Liu, Z. 2024. Initializing Models with Larger Ones. In *The Twelfth International Conference on Learning Representations*.

Yosinski, J.; Clune, J.; Bengio, Y.; and Lipson, H. 2014. How transferable are features in deep neural networks? In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N.; and Weinberger, K., eds., *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Yosinski, J.; Clune, J.; Nguyen, A.; Fuchs, T.; and Lipson, H. 2015. Understanding neural networks through deep visualization. In *International Conference on Machine Learning*, 2067–2075.

Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; and Vinyals, O. 2021. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM*, 64(3): 107–115.

Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; and He, Q. 2020. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1): 43–76.