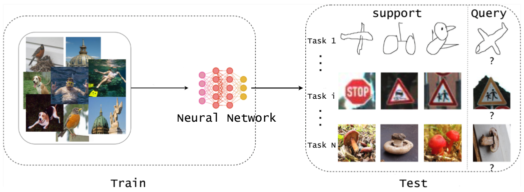
Channel Importance Matters in Few-shot Image Classification

Xu Luo, Jing Xu, Zenglin Xu

ICML International Conference On Machine Learning

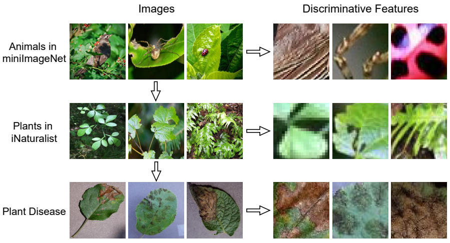
Background

Few-shot learning



Goal: train a visual model that can quickly *learn* new visual concept from a few examples.

Motivation



Three types of task distribution shifts:

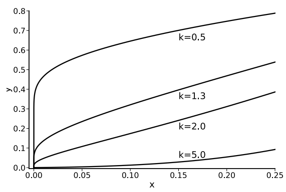
- **Category shift** (*mini*ImageNet, CIFAR)
- **Domain shift** (BSCD-FSL, MetaDataset)
- **Granularity shift** (Coarse-to-fine FSL)

Observation: Different tasks focus on different image features.
Guess: Task distribution shift leads to a biased focus of neural networks on image features.
Question: How to verify this guess?
Thinking: If such a bias is general, then there should be a feature transformation that can rectify this bias.

A Simple Channel-wise Feature Transformation

$$\phi_k(\lambda) = \begin{cases} \frac{1}{ln^k(\frac{1}{\lambda}+1)}, & \lambda > 0 \\ 0, & \lambda = 0 \end{cases}$$

We apply this function channel-wisely to image features only at test time.



Effect on 5-way 5-shot few-shot tasks:

Train \ Test	mini	CUB	Texture	TS	PlantD	ISIC	ESAT	Sketch	QDraw	Fungi
mini	-0.4	+2.3	+2.4	+1.4	+5.1	+1.0	+2.9	+2.6	+6.5	+3.3
CUB	+0.5	-0.2	+1.0	+2.4	+2.4	+2.3	+0.8	+1.0	+4.3	+1.1
Texture	+0.2	+0.1	-0.2	+0.1	+0.3	+0.2	+0.3	+0.1	+0.1	+0.0
TS	+2.2	+2.3	+1.6	+0.0	+3.5	+2.6	+2.5	+2.0	+4.5	+1.7
PlantD	+2.0	+2.1	+1.0	+0.6	+0.0	+1.0	+3.7	+0.6	+7.6	+1.6
ISIC	+0.1	+0.6	+0.3	-0.2	+2.0	-0.8	+1.0	+0.1	+0.8	+0.5
ESAT	+0.5	+0.2	+0.7	+0.5	+4.0	+0.0	-0.1	+1.5	+4.4	+0.6
Sketch	+2.4	+2.1	+2.6	+5.0	+4.2	+1.8	+6.9	-0.7	+2.0	+2.5
QDraw	+2.2	+0.9	+0.2	+1.2	+5.3	+1.3	+1.4	+1.8	-0.5	+1.4
Fungi	+3.1	+5.5	+3.6	+5.4	+3.6	+4.3	+6.4	+2.8	+5.7	-0.9

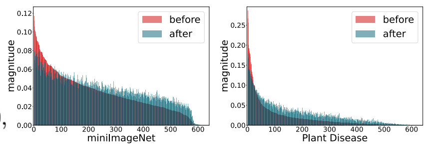
This transformation works only when task distribution shift from training to testing exists!

Conclusion & Question: We verify that as a result of task distribution shift, some kind of Bias of image features exists. Then How to characterize the bias? To answer this question, we need to further analyze this transformation.

Initial Analysis

Properties of the transformation:

$$\phi_k'(\lambda) > 0, \lim_{\lambda \to 0^+} \phi_k'(\lambda) = +\infty,$$
$$\exists t > 0, \quad s.t. \quad \forall \lambda \in (0,t), \phi_k''(\lambda) < 0,$$



Mean Magniture of Channels (**MMC**)

This transformation suppresses channels of large MMC and largely amplifies channels of small MMC.
MMC is important! Can we obtain the optimal or oracle MMC of any task?

Oracle MMC of Any Binary Task

Theorem: Two classes of image representations, each with mean and variances $\boldsymbol{\mu_1}, \boldsymbol{\sigma_1}$ and $\boldsymbol{\mu_2}, \boldsymbol{\sigma_2}$, respectively. Then the *oracle* MMC of the c-th channel ω_c should satisfy

$$\omega_c \propto \frac{|\mu_1^c - \mu_2^c|}{\sigma_1^c + \sigma_2^c}$$

The larger the mean difference, the smaller the variance, and the more important the channel.

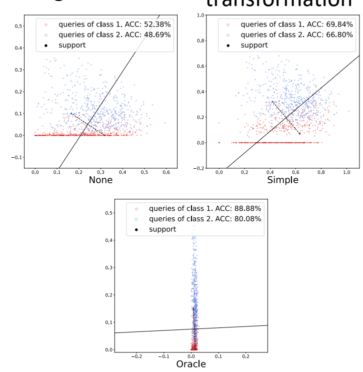
Empirical verification:

Algorithm	Classifier	Transformation	mini	CUB	Texture	TS	PlantD	ISIC	ESAT	Sketch	QDraw	Fungi	Avg
PN	NCC	None	90.5	80.6	80.6	85.1	89.2	65.7	86.5	71.9	82.4	74.6	80.7
		Simple	91.3	82.4	83.1	85.8	93.0	68.6	89.2	75.2	85.1	77.2	83.1
		Oracle	**93.1**	**88.7**	**87.2**	**92.4**	**95.6**	**69.1**	**91.5**	**81.2**	**89.4**	**88.4**	**87.7**
S2M2	LC	None	94.0	87.1	85.7	88.7	95.0	68.7	93.5	78.7	85.5	82.8	86.0
		Simple	94.4	88.3	87.3	91.2	96.4	72.2	93.8	81.0	89.2	84.5	87.8
		Oracle	**96.3**	**94.0**	**90.7**	**96.1**	**98.3**	**72.6**	**95.2**	**87.0**	**93.0**	**93.3**	**91.7**

Visualization:

Original feature

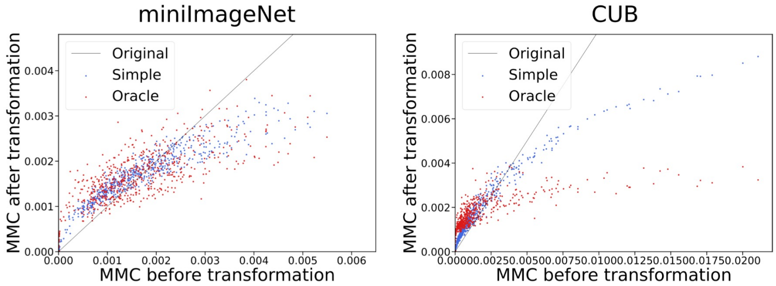
Simple transformation



Oracle feature

- The two axis are two selected channels from feature representations learned by ProtoNet. We here show a binary 1-shot task on these two channels using the ProtoNet classifier head. The black line is the classification boundary.
- The rescaling of channels made by simple and oracle transformation helps highlight the most discriminative channel (y-axis).

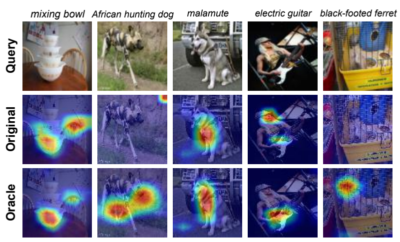
Analysis of Channel Bias of Visual Representations



Comparisons of MMC. A point is a channel; x-axis is the original MMC, y-axis is the transformed MMC.

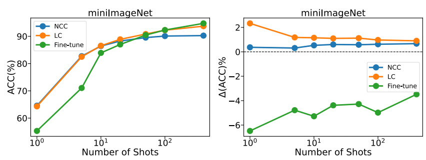
Conclusions:

- Neural networks are overconfident in previously learned channel emphasis.
- The channel bias problem diminishes as task distribution shift lessens.
- The simple transformation pushes channel emphasis towards the optimal ones.



- The channel bias problem distracts the neural network from new objects.

Shot Analysis



- The channel bias problem requires more attention in few-shot setting, while simple fine-tuning can help address this problem in many-shot setting.
- Logistic regression can alleviate the channel bias problem to some extent in many shot setting.

Code is publicly available at https://github.com/Frankluox/Channel_Importance_FSL