State-of-art deep learning technologies and their application to air-shower reconstruction

Vladimir Sotnikov

JetBrains Research

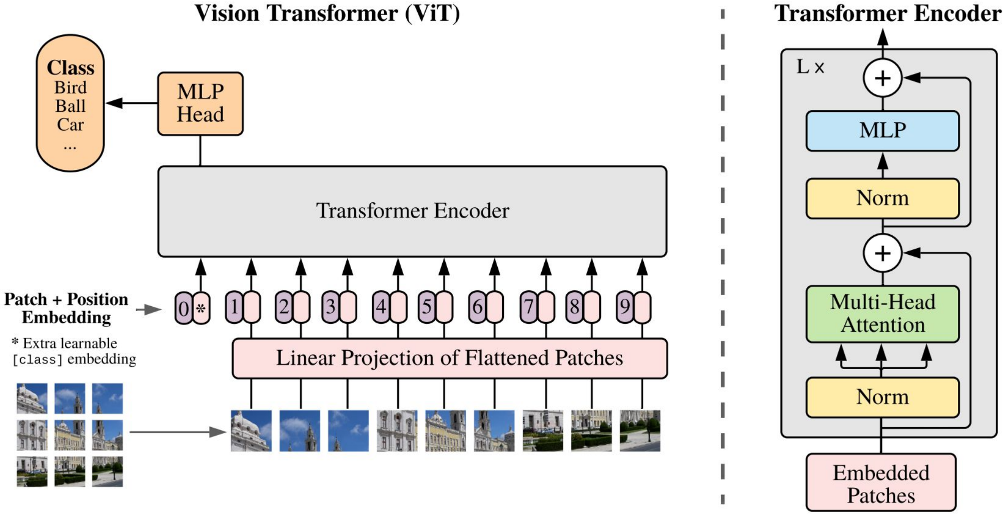
February 2, 2022

About me / Outline

- Deep learning developer at JB Research and JB Computational Arts
- 5 years of experience in commercial computer vision (CV)
- Started working with air showers about a year ago
- Analyzing KASCADE archive data using deep neural networks
- In this talk I will try to highlight some of the most promising deep learning techniques that could be applied to air showers

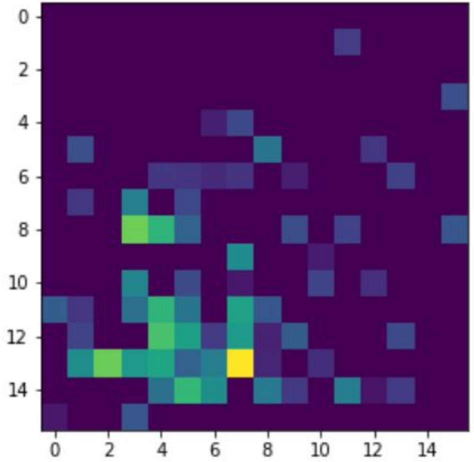
Vision Transformers (ViT) and Attention MLPs

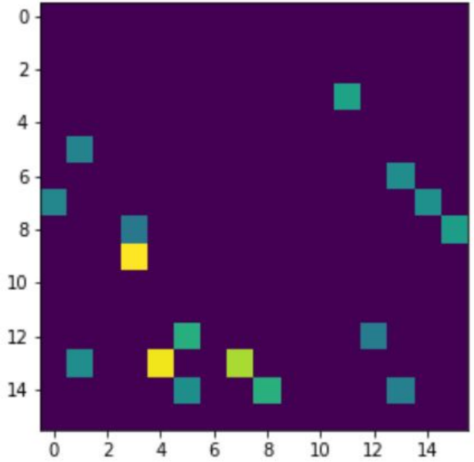
- Just like the original natural language processing (NLP) Transformers, ViTs are only using self-attention and feedforward layers
- Input image is being represented as a set of fixed-size patches
- Attention mechanism, combined with position embedding for each patch, aggregates information across locations



Vision Transformers (ViT) and Attention MLPs

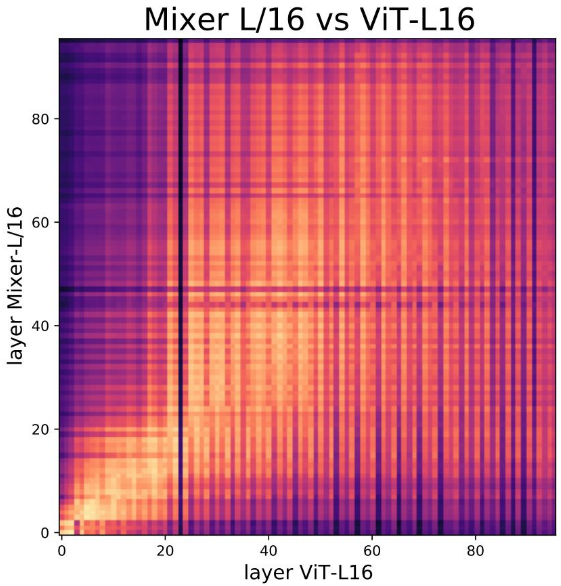
- Unlike CNNs, ViTs and MLPs are using spatial information
- One of the key rationales for CNNs (back in AlexNet, LeNet time) was the locality of pixel dependencies which is not always the case with air showers

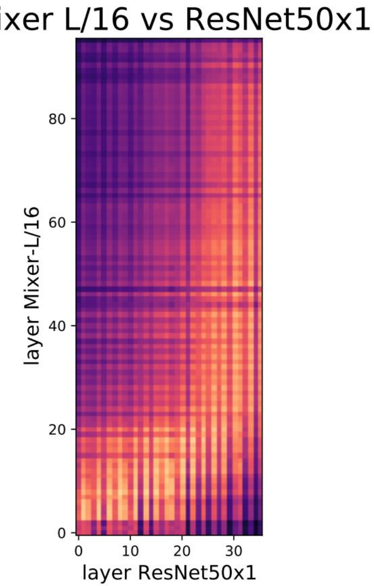




ViTs and CNNs - perception differences

- ViT incorporates more global information than ResNet at lower layers
- Skip connections in ViT are even more influential than in ResNets
- ViTs internal representations are similar to MLP-Mixer despite the latter not using attention





Unsupervised pre-training

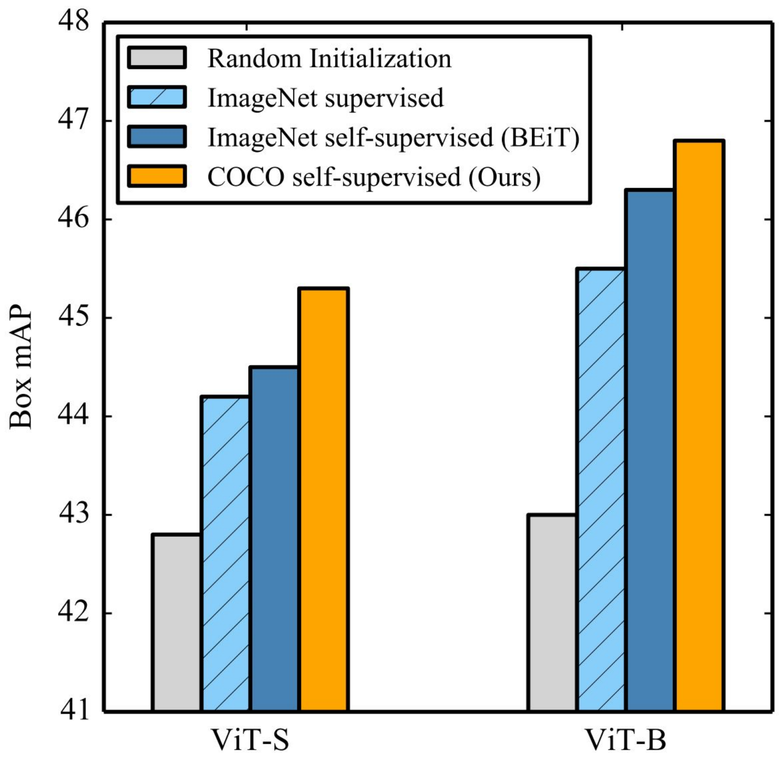
- Becoming increasingly popular since the GPT-1 release
- First was applied to NLP tasks, now expanding to CV as well
- Relevant when unlabeled data are abundant while labeled data are scarce
- For air showers, it allows to employ experimental data

Unsupervised pre-training

- Becoming increasingly popular since the GPT-1 release
- First was applied to NLP tasks, now expanding to CV as well
- Relevant when unlabeled data are abundant while labeled data are scarce
- For air showers, it allows to employ experimental data
- Potentially increases robustness of the model and/or decreases the required amount of training data
- Application to air showers is nontrivial

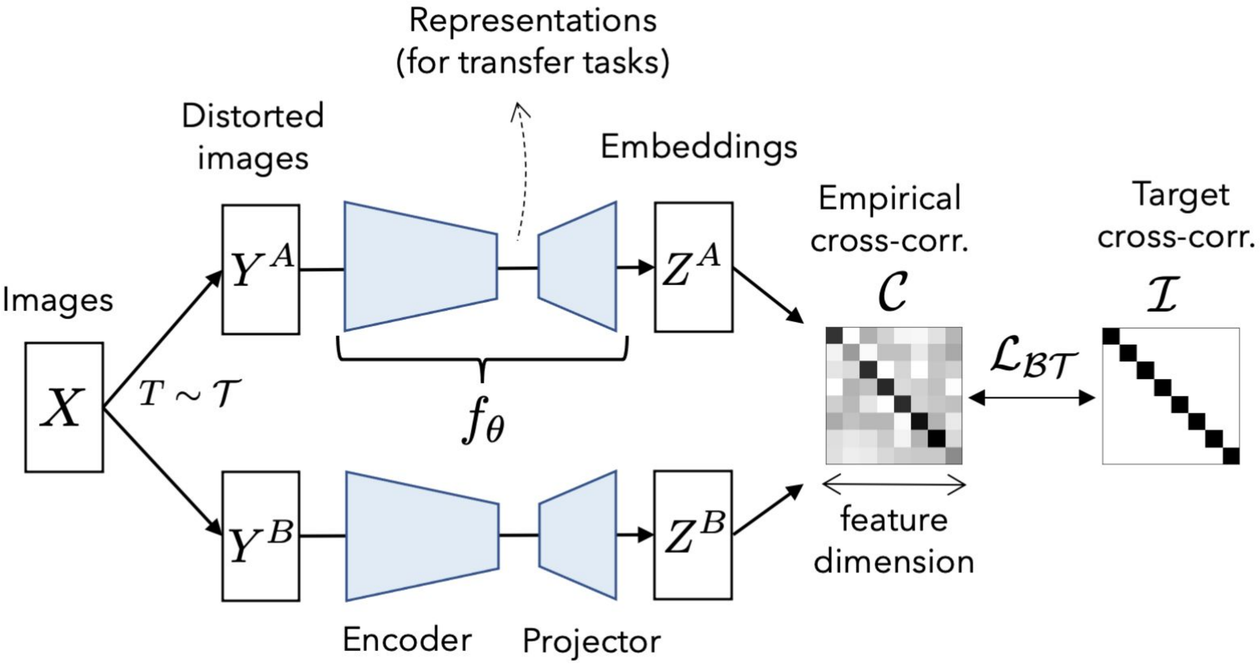
Unsupervised pre-training

- Interestingly, there are also some evidences that pre-training on targeted (i.e. labeled) dataset also improves model performance



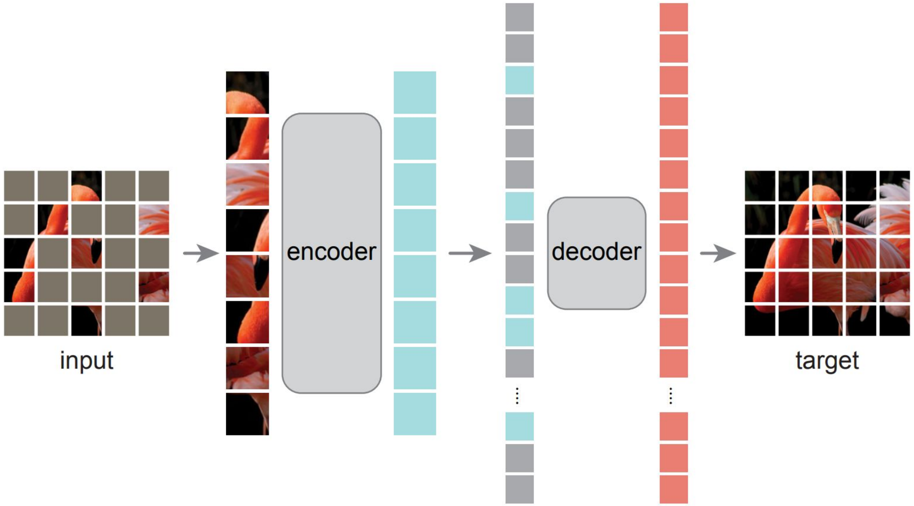
Unsupervised pre-training - Barlow Twins

- Requires a pair of identical neural networks
- Twin neural networks compute embedding for the same (but differently augmented) images
- Compared to contrastive loss, doesn't require lots of negative samples per batch or low-dimensional embeddings
- Tricky to apply to air-showers - very few possible augmentations



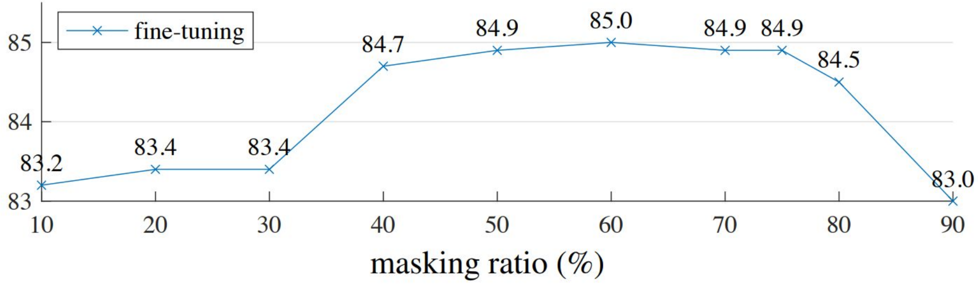
Unsupervised pre-training - Masked Autoencoders (MAE)

- Random patches of the image are being masked (at a very high proportion)
- Visible patches are feed to the encoder
- Lightweight decoder reconstructs the whole image from patches and mask tokens
- After pre-training, encoder is being used on unmasked images, decoder is discarded
- Requires relatively high-dimensional redundant input



Unsupervised pre-training - Masked Autoencoders (MAE)

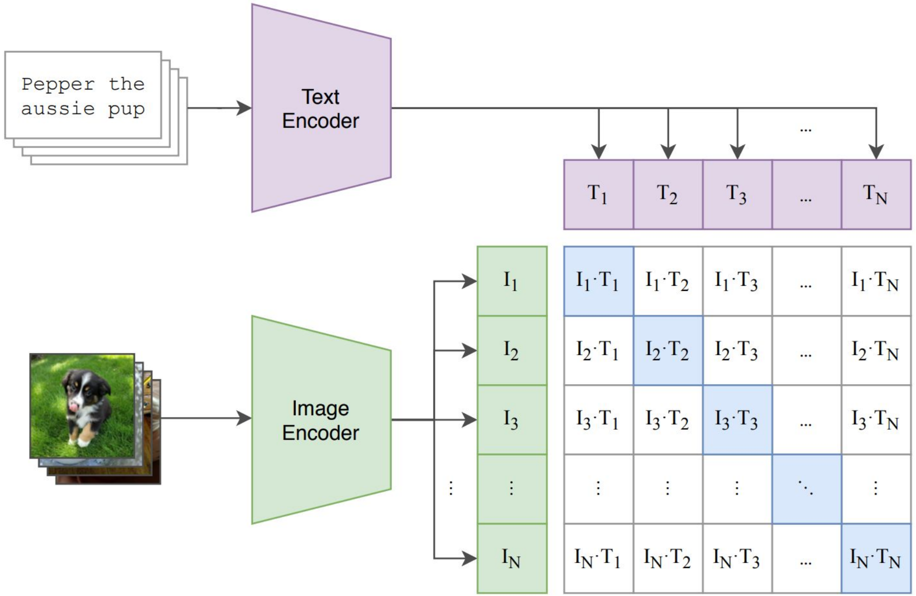
- Random patches of the image are being masked (at a very high proportion)
- Visible patches are feed to the encoder
- Lightweight decoder reconstructs the whole image from patches and mask tokens
- After pre-training, encoder is being used on unmasked images, decoder is discarded
- Requires relatively high-dimensional redundant input



Unsupervised pre-training - other approaches

These approaches seem to be fundamentally incompatible with the air shower domain and won't be covered in detail:

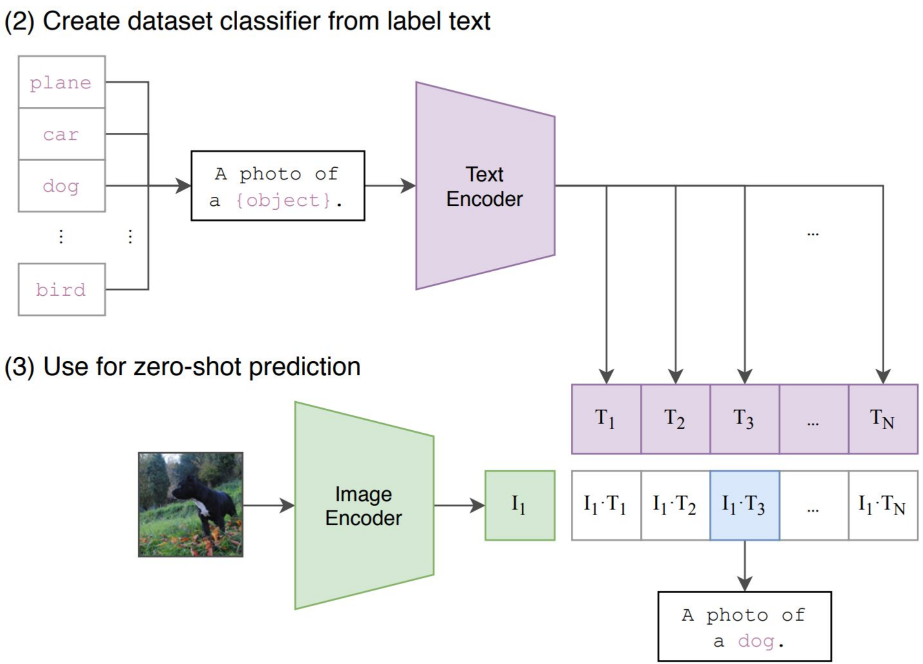
- Weakly Supervised Pre-Training - uses noisy semantic learning signal (hashtags) associated with the data
- Contrastive Language-Image Pre-Training - probably one of the most revolutionary zero-shot pre-training approaches, which, however, requires natural language supervision.



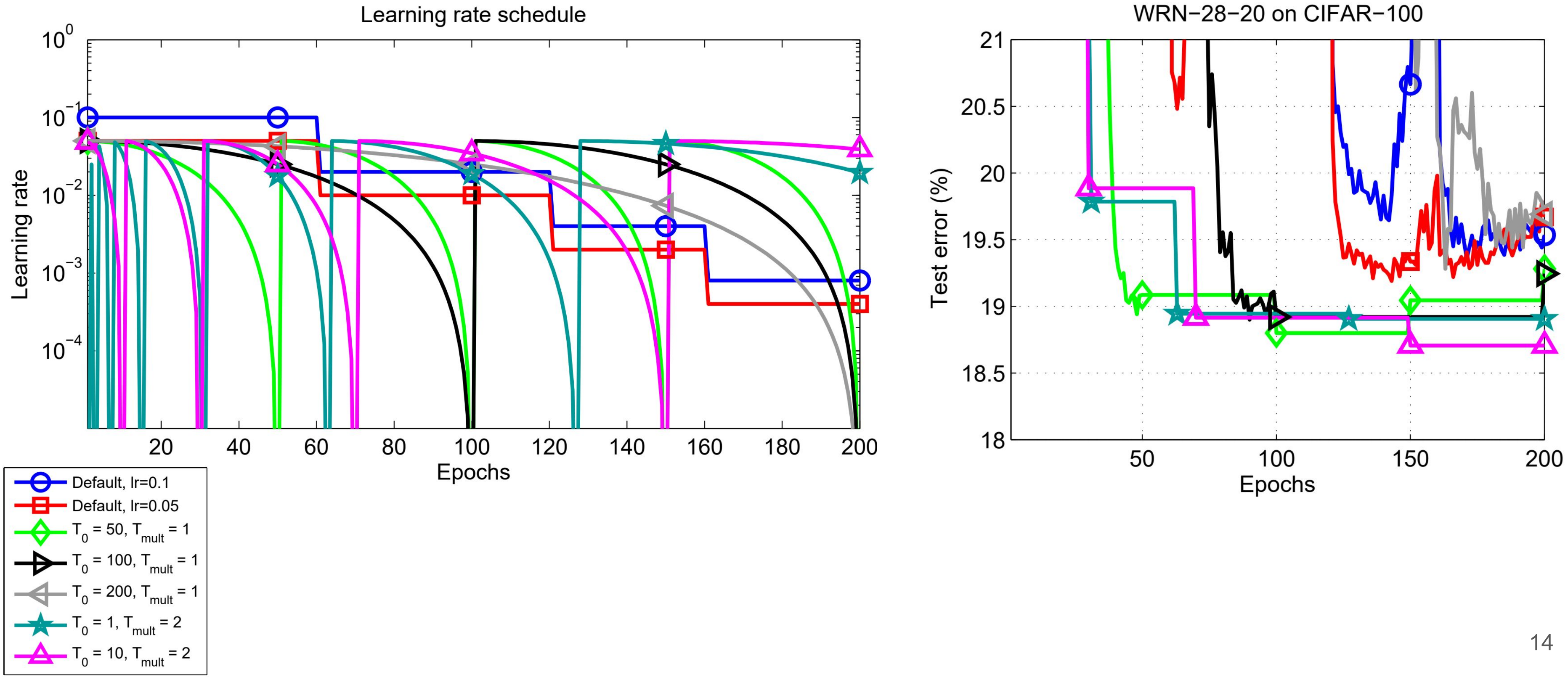
Unsupervised pre-training - other approaches

These approaches seem to be fundamentally incompatible with the air shower domain and won't be covered in detail:

- Weakly Supervised Pre-Training - uses noisy semantic learning signal (hashtags) associated with the data
- Contrastive Language-Image Pre-Training - probably one of the most revolutionary zero-shot pre-training approaches, which, however, requires natural language supervision.

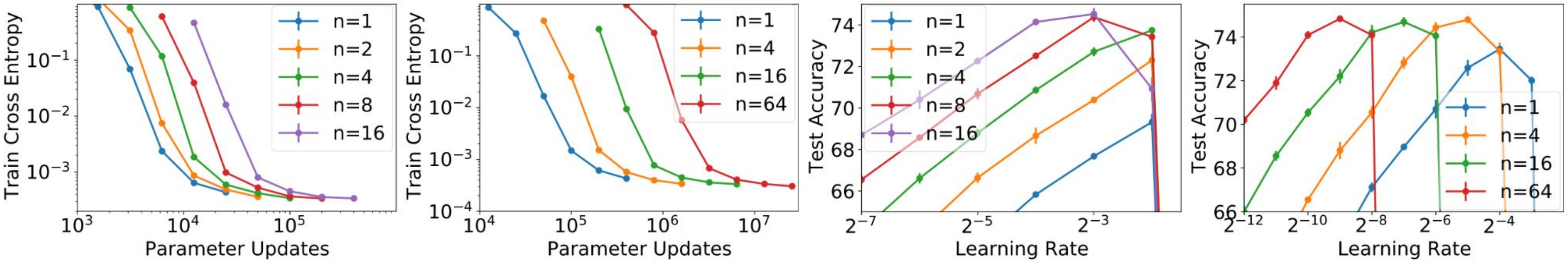


Cosine LR schedule with warm restarts



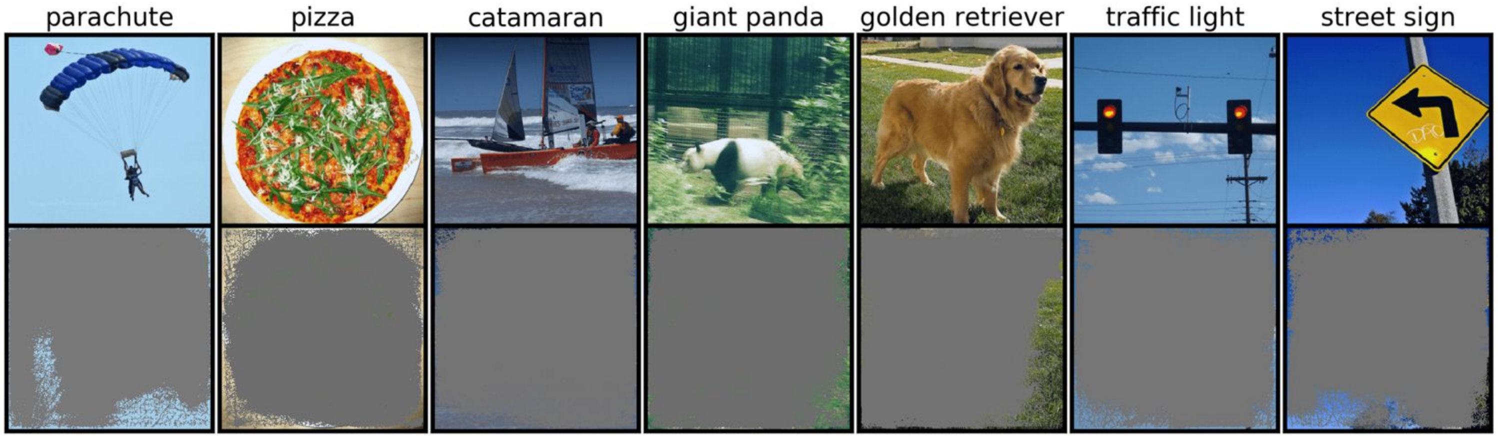
Multiple Augmentation Samples Per Image Decreases Test Error

- Model shows better performance with fixed batch size as the augmentation multiplicity rises (and as amount of unique samples goes down)
- Model shows better performance with fixed amount of unique samples as the augmentation multiplicity rises (and as batch size grows)
- Bigger batch size requires higher LR

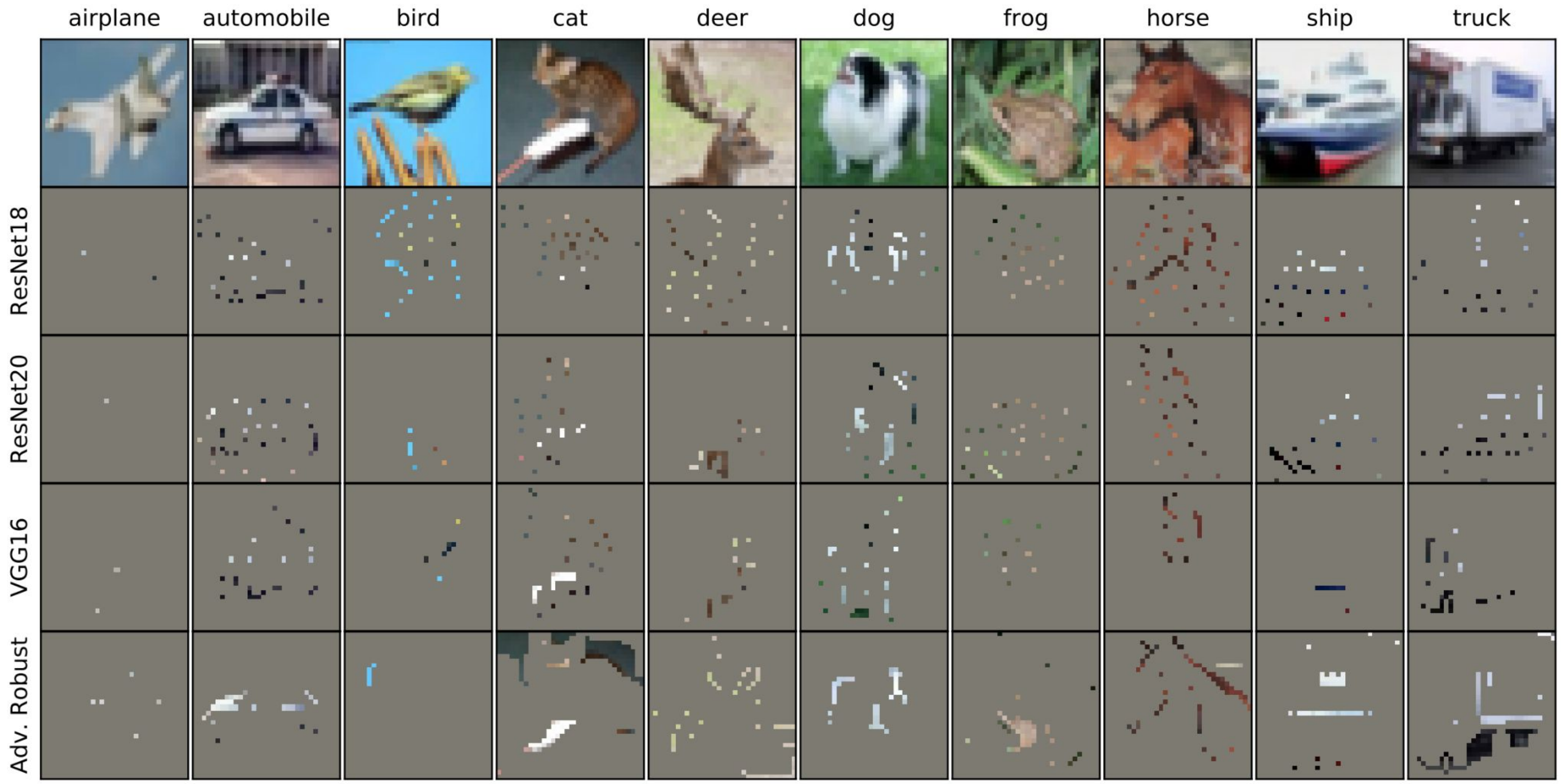


Overinterpretation reveals image classification model pathologies

- In some cases high accuracy of the model could be explained by overinterpreting unintended nonsensical patterns (90%+ confidence correct validation samples below)



Overinterpretation reveals image classification model pathologies

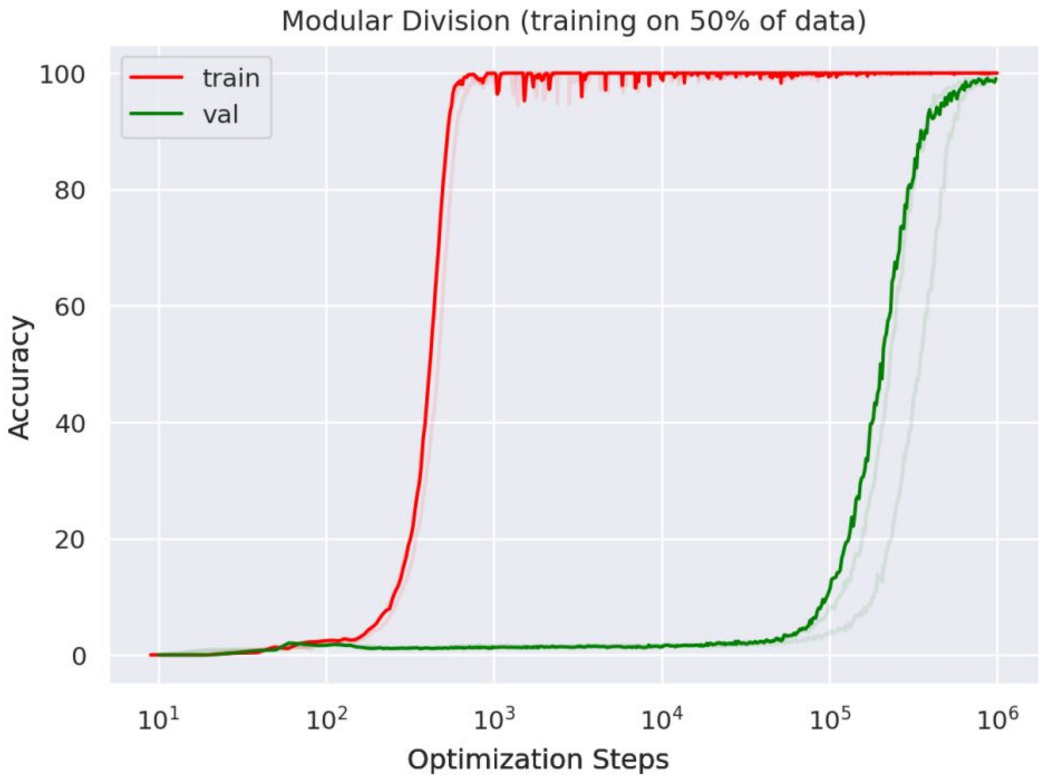


Overinterpretation reveals image classification model pathologies

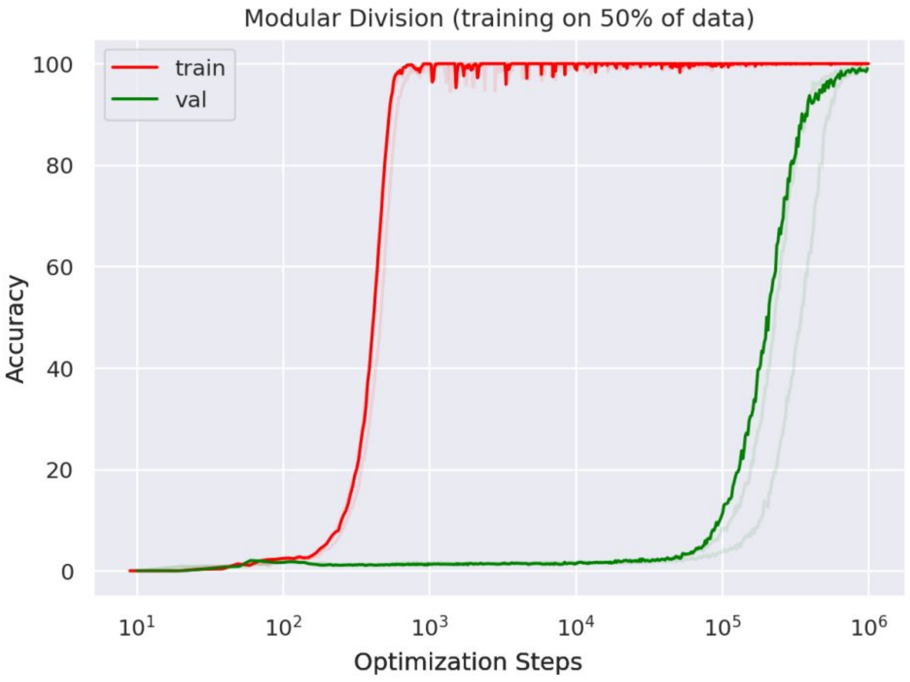
- Easy to implement alternative sanity check: freezing the model weights + adding trainable binary classification head
- High accuracy on MC/Exp classification will indicate that the model latent space preserves information about discrepancies between datasets

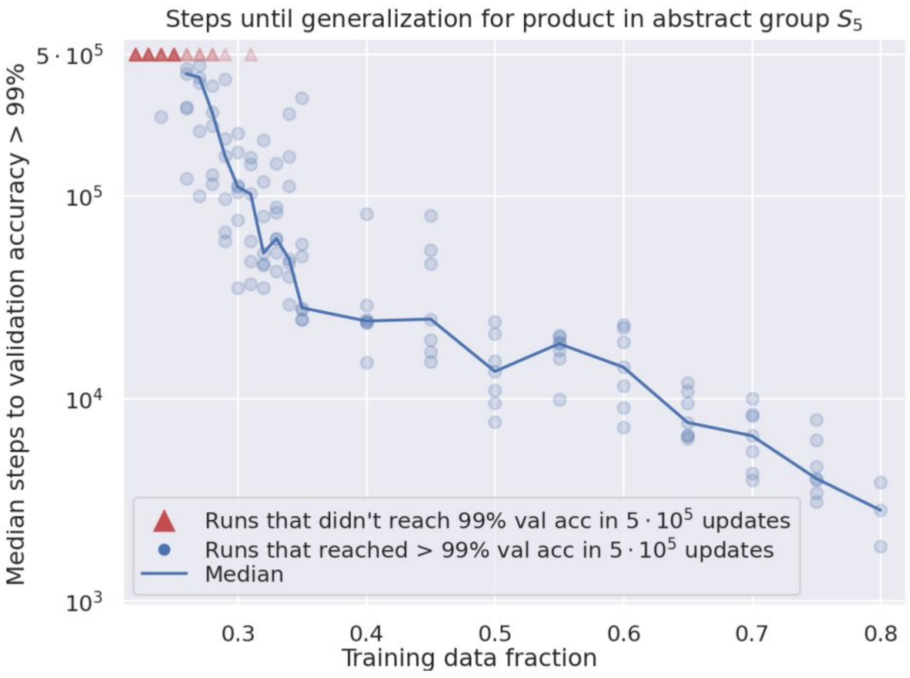
Generalization Beyond Overfitting

- After the memorization of training samples, validation accuracy sometimes suddenly begins to increase toward perfect generalization
- This phenomenon occurs under various circumstances but only with synthesized datasets



Generalization Beyond Overfitting





Conclusions

- Non-convolutional architectures are becoming increasingly popular
- While being a very powerful technique, self-supervised pre-training is nontrivial to apply to air showers
- Some of the covered methods could improve model efficiency and/or interpretability
- Feel free to reach us out:
  - astroparticle@jetbrains.com
  - https://research.jetbrains.org/groups/astroparticle-physics/