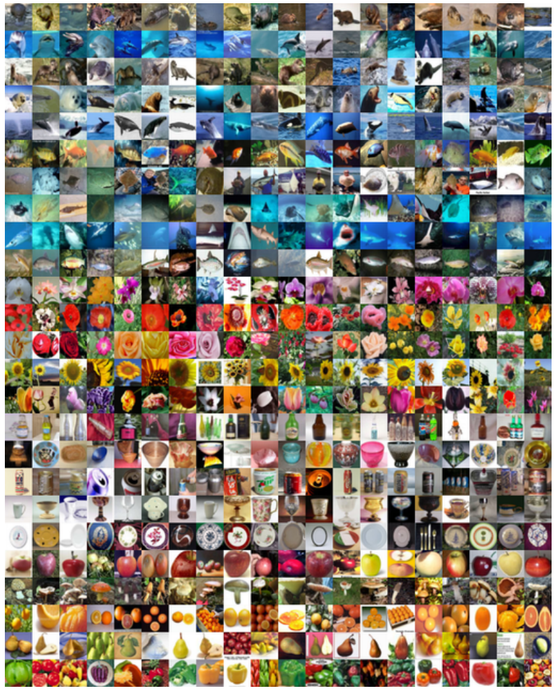


CIFAR100

Deep Learning - Part B

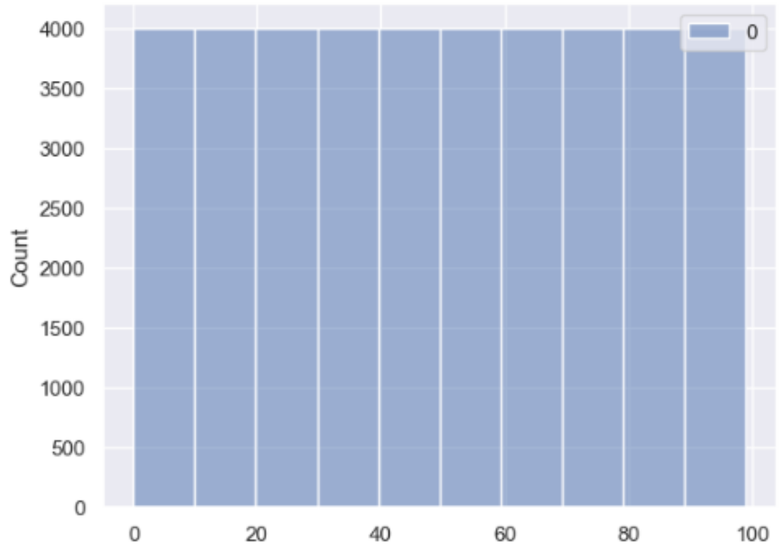
by

Kenneth Chen (2100072)

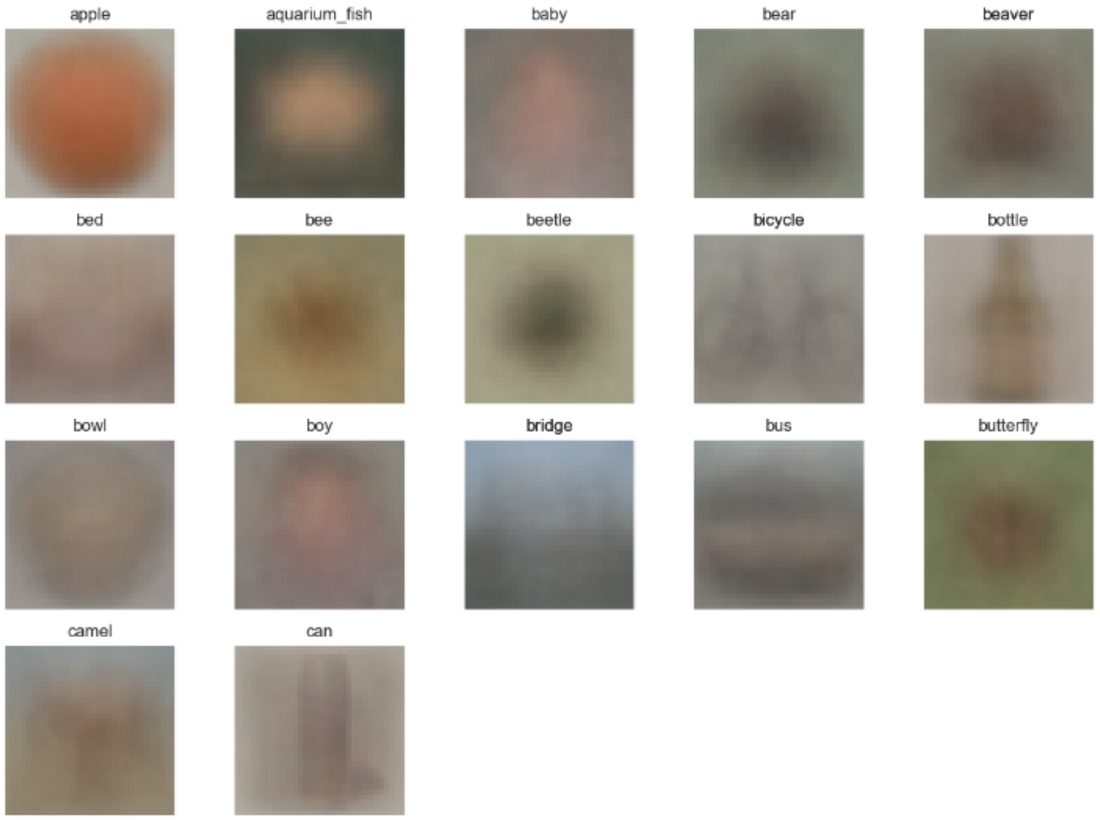
1. EXPLORATORY DATA ANALYSIS
2. DATA AUGMENTATION
3. MODELING
4. IMPROVEMENT
5. FINAL ANALYSIS

Exploratory Data Analysis

The basics



no class imbalance



average of first 20 fine classes

Exploratory Data Analysis

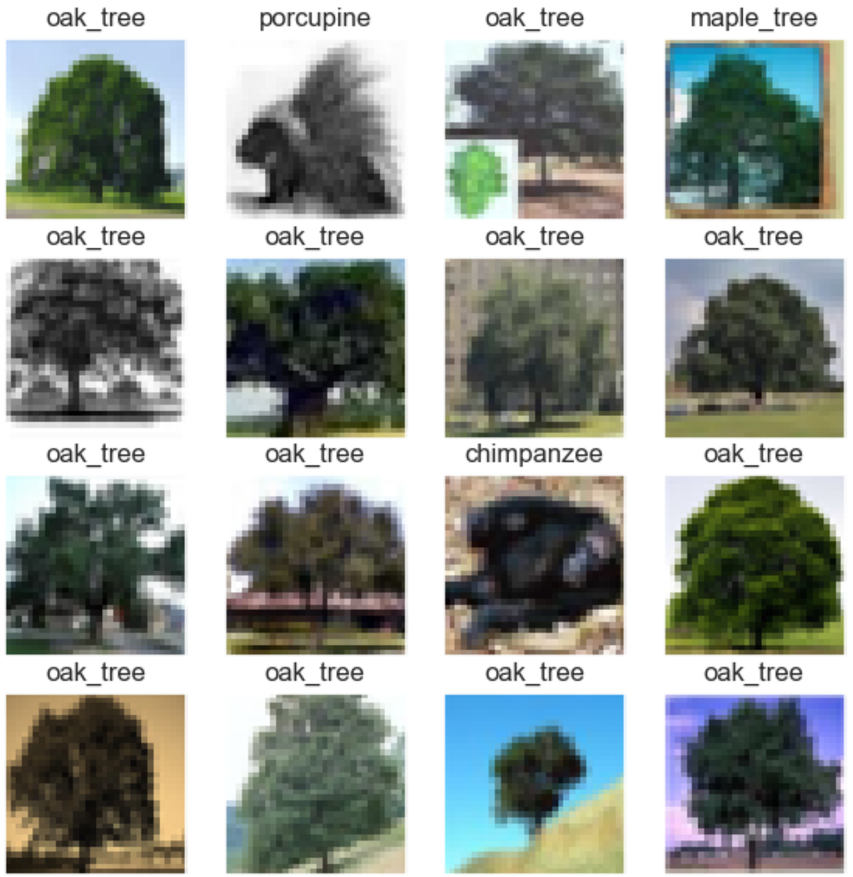
Outlier analysis

ResNet-S
Feature Extractor

Last two layers are removed to give feature embedding

Embedding vector passed to **TSNE** (n_components = 4)

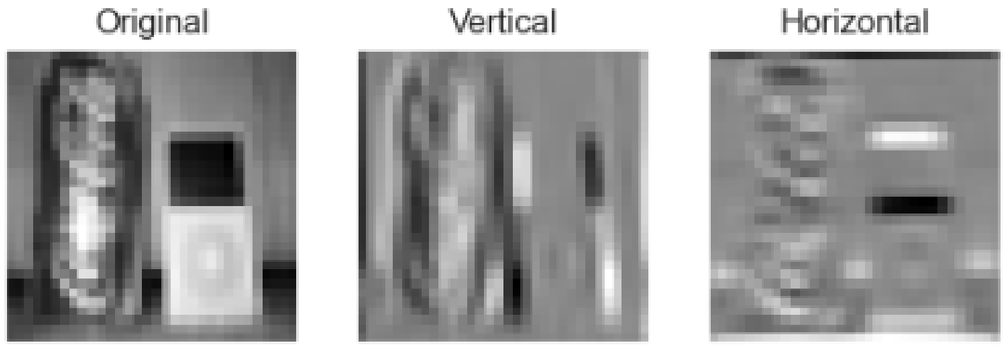
Class-wise distance from class mean



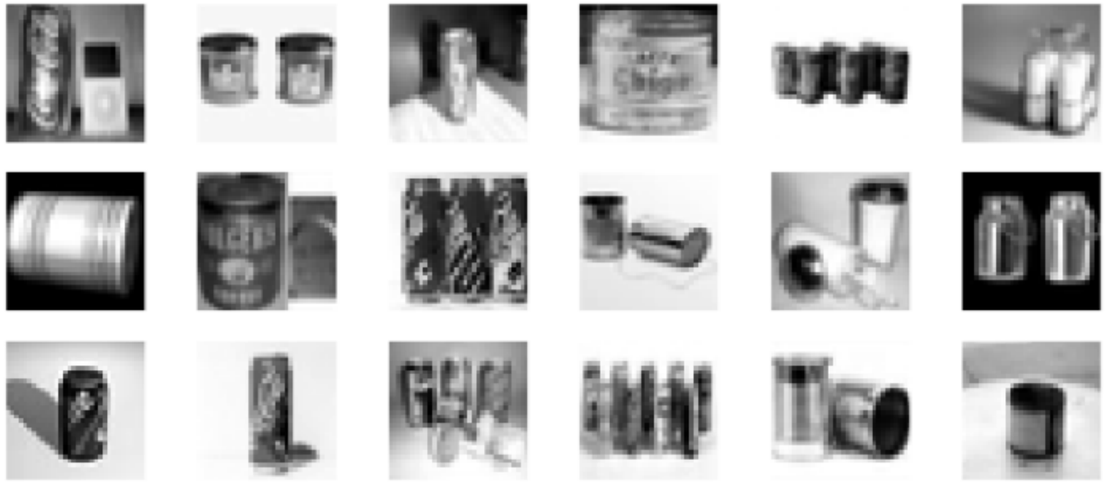
Exploratory Data Analysis

Rotation analysis

Hand-crafted
convolution kernel
to **extract** vertical and
horizontal lines



Engineered and
designed algorithm to
detect **rotated images**



What Model to use?

1

MINI ALEXNET

A smaller AlexNet.
Train: 0.20955
Val: 0.3085

2

VGG15

Custom VGG Architecture.
Train: 0.8854
Val: 0.6427

3

RESNET50

50-Layer Deep Residual Network.
Train: 0.6668
Val: 0.6601

4

CVT

Convolutional Vision Transformers
Train: 0.1583
Val: 0.1999

5

EFFICIENTNET B0

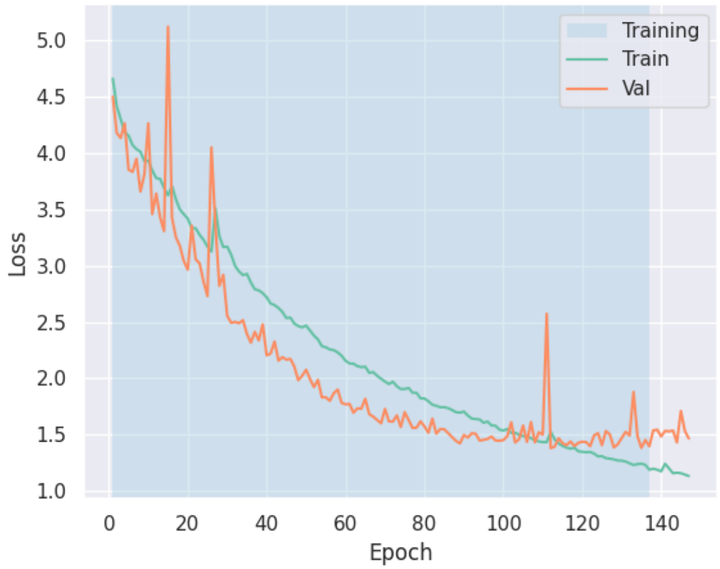
MBConv + Squeeze and Excitation.
Train: 0.768
Val:0.478

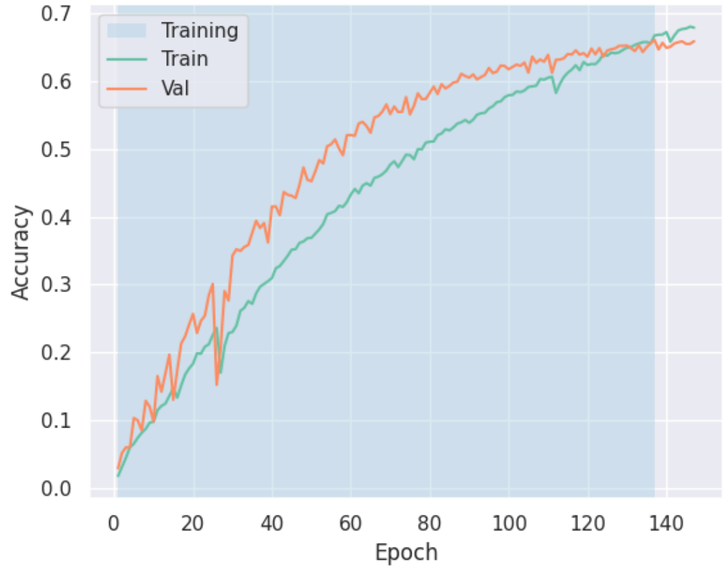
What Model to use?

RESNET50

50-Layer Deep Residual Network

~25 million parameters

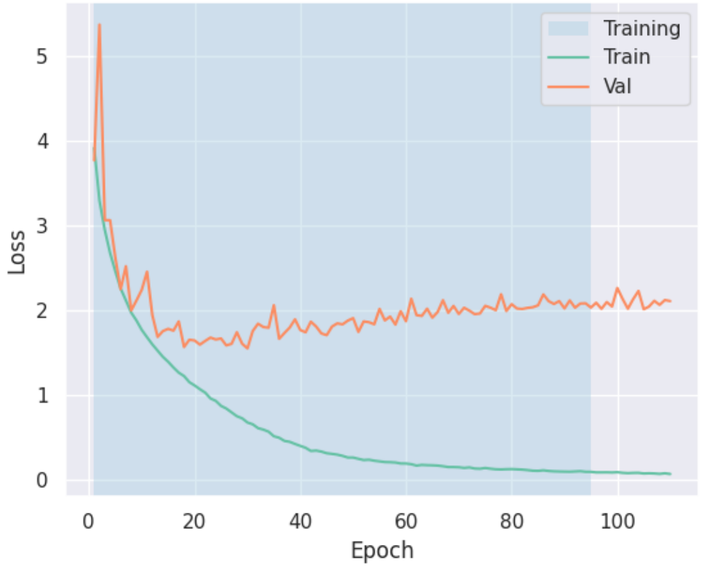


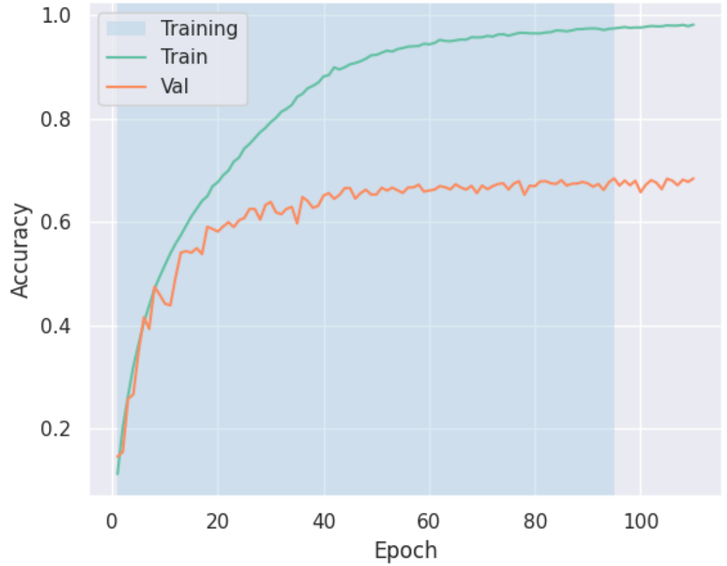


Let's proceed with Residual Networks

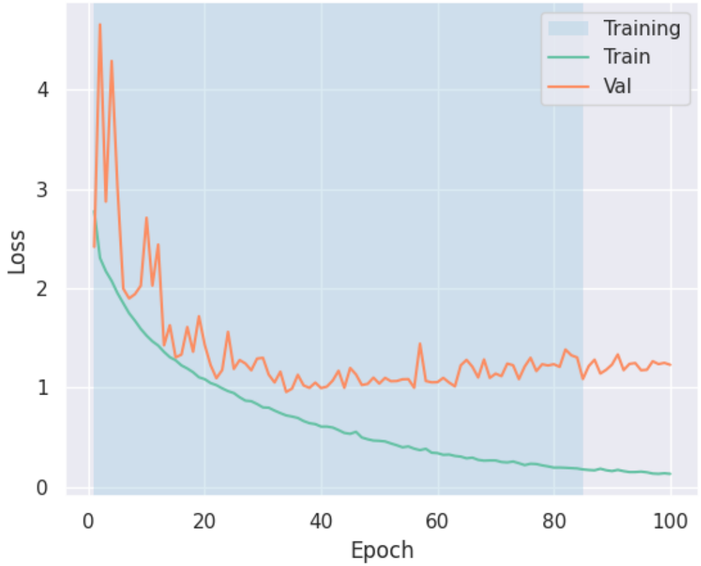
MODEL	DESCRIPTION	TRAIN ACCURACY	VALIDATION ACCURACY
ResNet50	Simple ResNet 50	66.68	66.01
ResNeXt 29x4d	Combination of VGG, ResNet and Inception.	97.41	68.41
Wide ResNet	ResNet with larger expansion factor	96.13	59.91

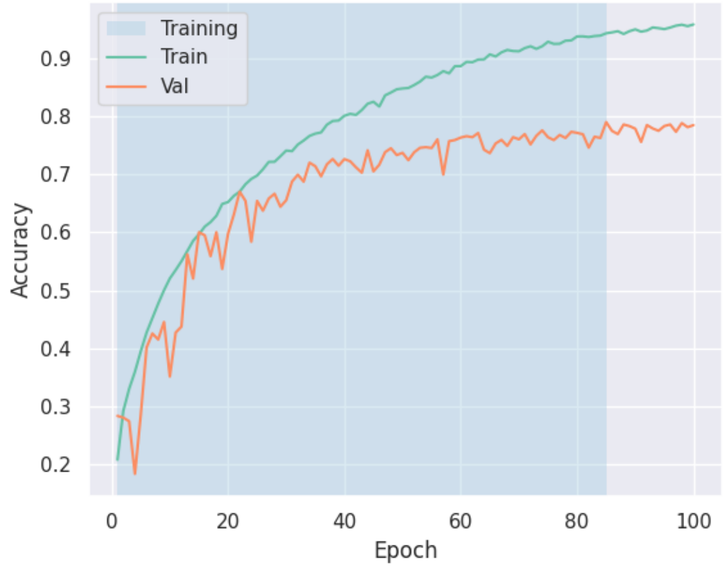
Fine labels. Val: 0.68





Coarse labels. Val: 0.76



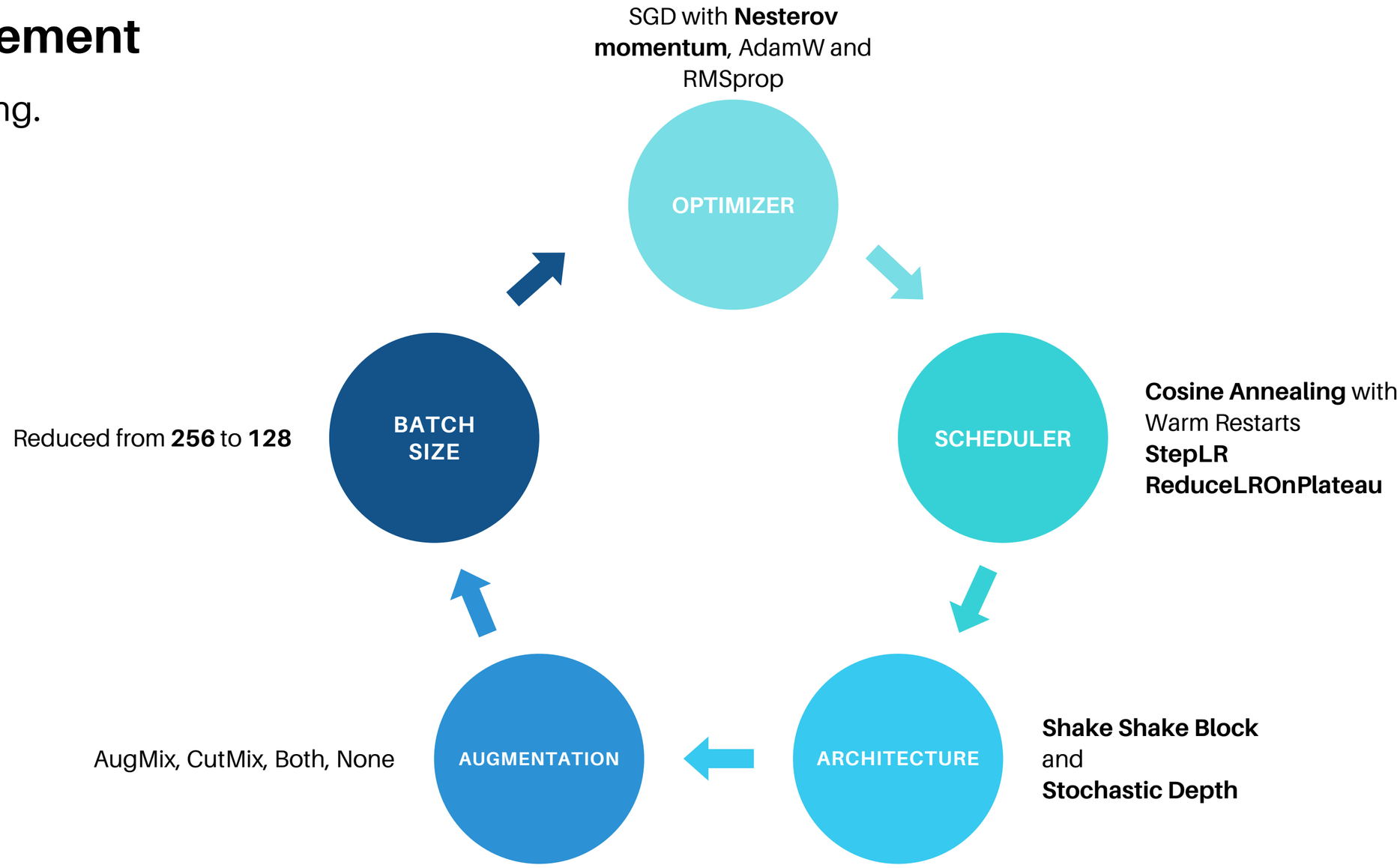


ResNeXt 29x4d

I discovered that simply mapping the fine output to coarse label gives a **validation score of 0.76**

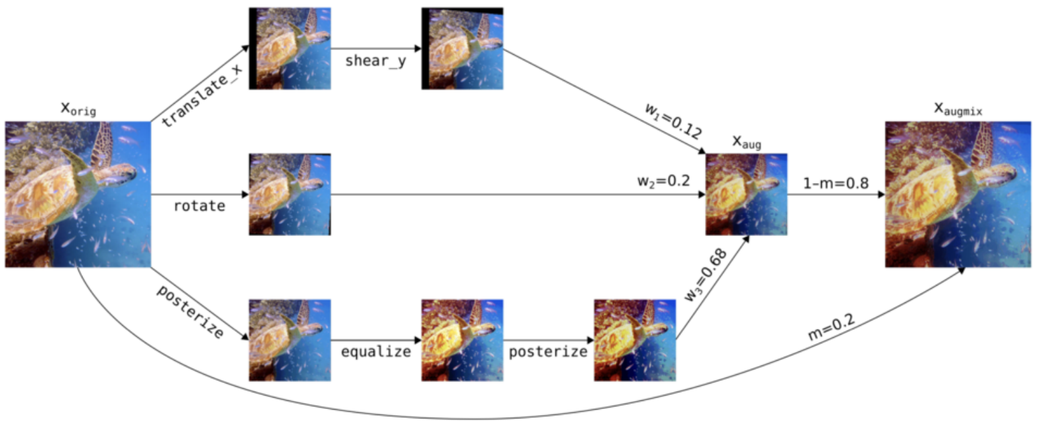
Model Improvement

Tuning and improving.

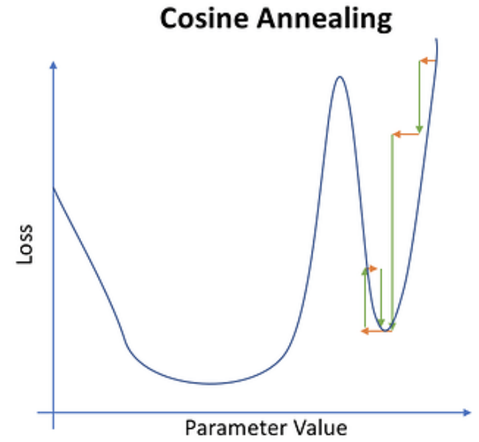


Model Improvement

Tuning and improving.



AUGMENTATION



Each step ( ) changing size (←)with cosine annealing, but not general and gets ‘stuck’.

SCHEDULER

Final Model

Trained and tuned on **4x NVIDIA RTX 3090s** for **10 hours**

1

MODEL

ShakeDrop
ResNeXt 29x4d
~75 million parameters

2

OPTIMIZER

SGD with
Nesterov momentum

3

SCHEDULER

Cosine Annealing with
Warm Restarts

4

AUGMENTATION

AugMix + CutMix

5

RESULTS

Train: **77.4%**
Val: **75.8%**
Fine Test: **75.6%**
Coarse Test: **84.4%**

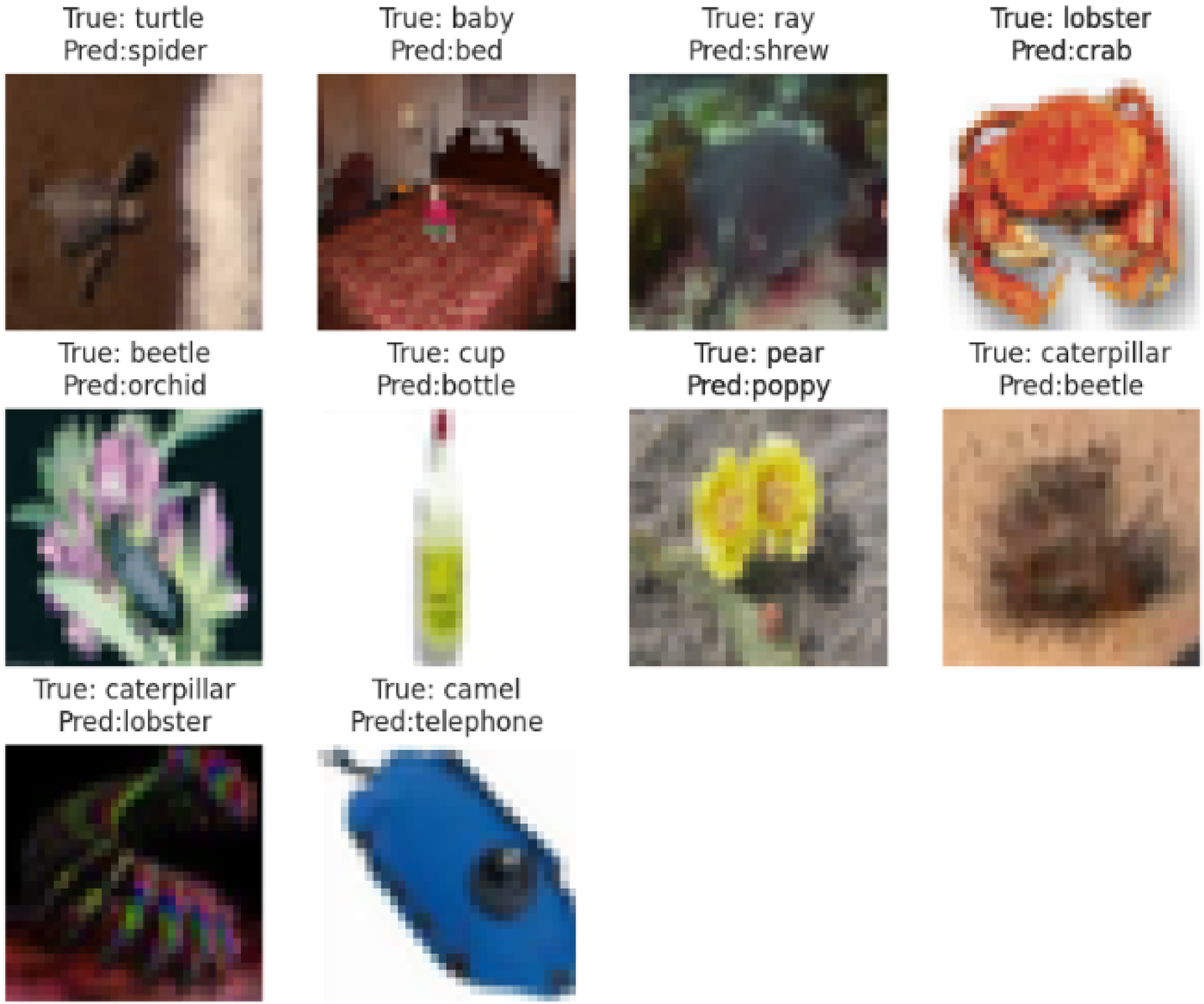
The best classes

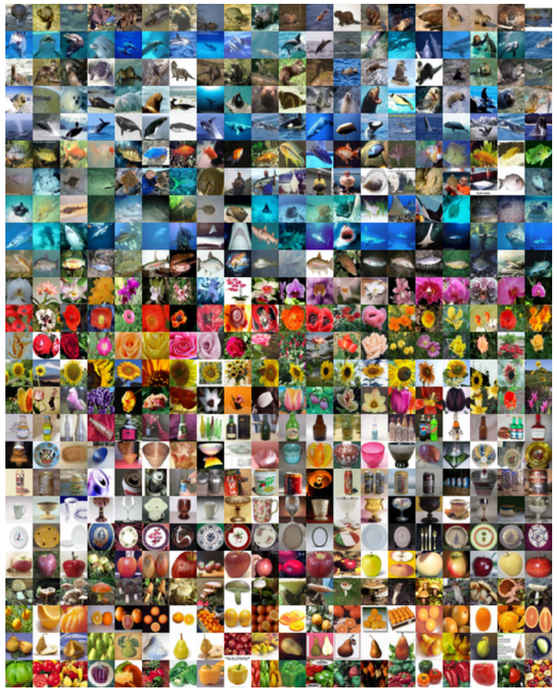
Class	Accuracy
Skunk	0.98
Motorcycle	0.96
Road	0.96
Wardrobe	0.94
Orange	0.94

The worst classes

Class	Accuracy
Otter	0.46
Bowl	0.47
Girl	0.49
Willow Tree	0.52
Boy	0.54

Which images was the model most wrong on?





Thank you

References

- Adam, A. (2018) Early Stopping and its Faults, Early Stopping and its Faults. Available at: https://alexadam.ca/2018/08/03/early-stopping/ (Accessed: 25 October 2022).
- Burnham, K.P. and Anderson, D.R. (2010) Model selection and multimodel inference: a practical information-theoretic approach. 2. ed. New York, NY: Springer.
- Dosovitskiy, A. et al. (2020) 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale'. Available at: https://doi.org/10.48550/ARXIV.2010.11929.
- Gastaldi, X. (2017) 'Shake-Shake regularization'. arXiv. Available at: http://arxiv.org/abs/1705.07485 (Accessed: 22 November 2022).
- Hendrycks, D. et al. (2020) 'AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty'. arXiv. Available at: http://arxiv.org/abs/1912.02781 (Accessed: 22 November 2022).
- Huang, G. et al. (2016) 'Deep Networks with Stochastic Depth'. arXiv. Available at: http://arxiv.org/abs/1603.09382 (Accessed: 22 November 2022).
- Loshchilov, I. and Hutter, F. (2017) 'SGDR: Stochastic Gradient Descent with Warm Restarts'. arXiv. Available at: http://arxiv.org/abs/1608.03983 (Accessed: 22 November 2022).
- Loshchilov, I. and Hutter, F. (2019) 'Decoupled Weight Decay Regularization'. arXiv. Available at: http://arxiv.org/abs/1711.05101 (Accessed: 22 November 2022).
- Wu, H. et al. (2021) 'CvT: Introducing Convolutions to Vision Transformers'. Available at: https://doi.org/10.48550/ARXIV.2103.15808.
- Xie, S. et al. (2017) 'Aggregated Residual Transformations for Deep Neural Networks'. arXiv. Available at: http://arxiv.org/abs/1611.05431 (Accessed: 22 November 2022).
- Yamada, Y. et al. (2019) 'ShakeDrop Regularization for Deep Residual Learning', IEEE Access, 7, pp. 186126–186136. Available at: https://doi.org/10.1109/ACCESS.2019.2960566.
- Yun, S. et al. (2019) 'CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features'. arXiv. Available at: http://arxiv.org/abs/1905.04899 (Accessed: 22 November 2022).