EXPERIMENTS WITH IMAGE AUGMENTATION FOR **CLASSIFICATION** AND SEGMENTATION PART 1

V. Martinek

OVERVIEW

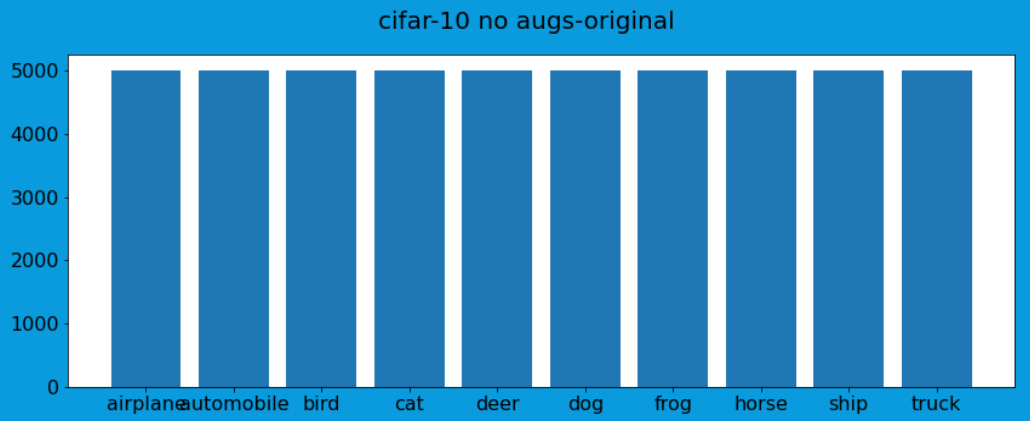
- My own experiments, not a research paper
- Goal – answer some questions:
  - Are augmentations always good?
  - Which of them are the best?
  - Are they dataset dependent?
  - How to compare better? (For part 2)
- 20 augmentations
  - Geometric
  - Colors
  - Erasing
  - Filters
- No augmentation % experiments , combinations etc…

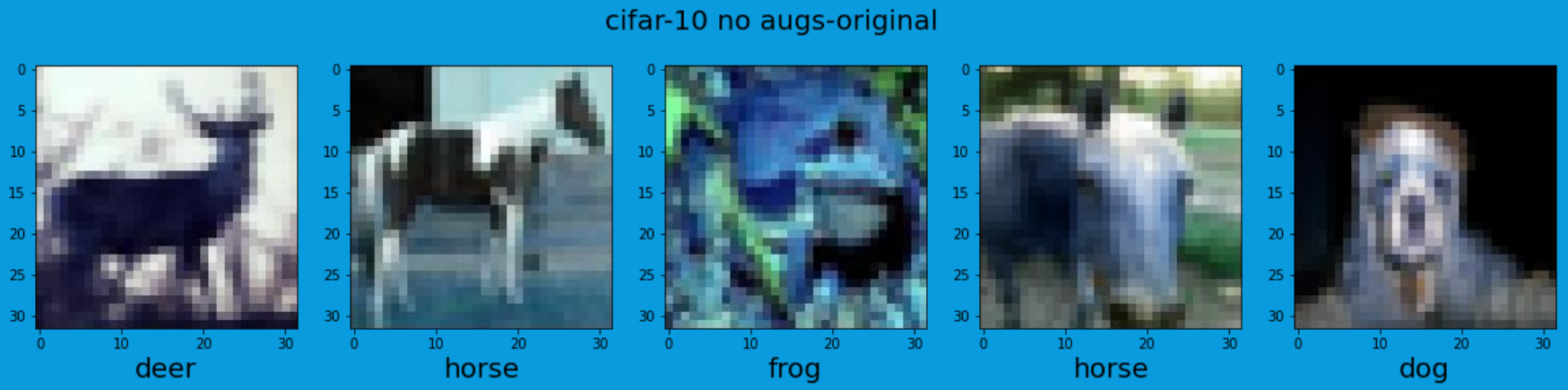
OVERVIEW

- Datasets -> preprocessing and augmentations -> models -> metrics -> good?
- 3 datasets
- 3 models
- Preprocessing
- Augmentations
- Training for 20 epochs
- Metrics measured on test set
- Dataset split based on metadata (hopefully less data leakage)

DATASETS – CIFAR-10

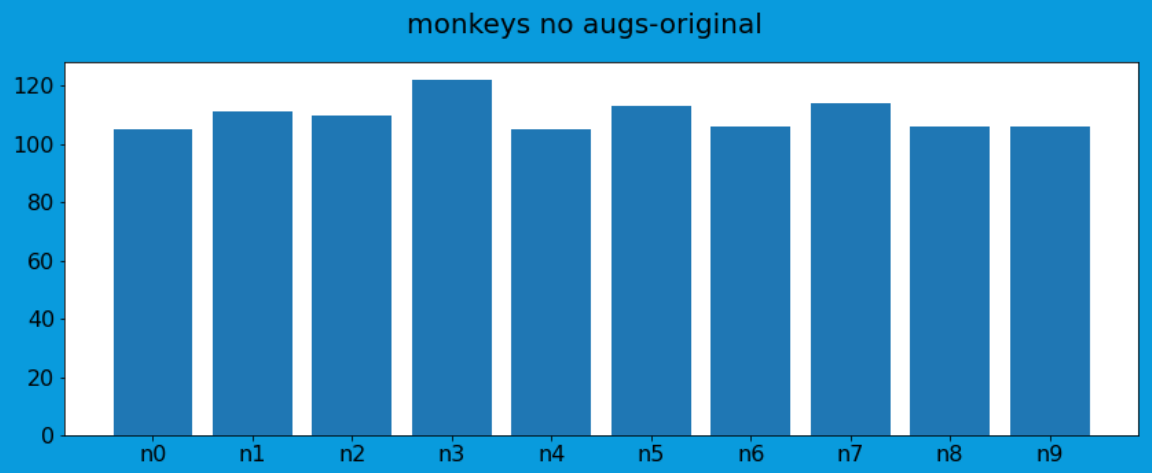
- 10 classes
- 5000 images of each class
- consistent resolution

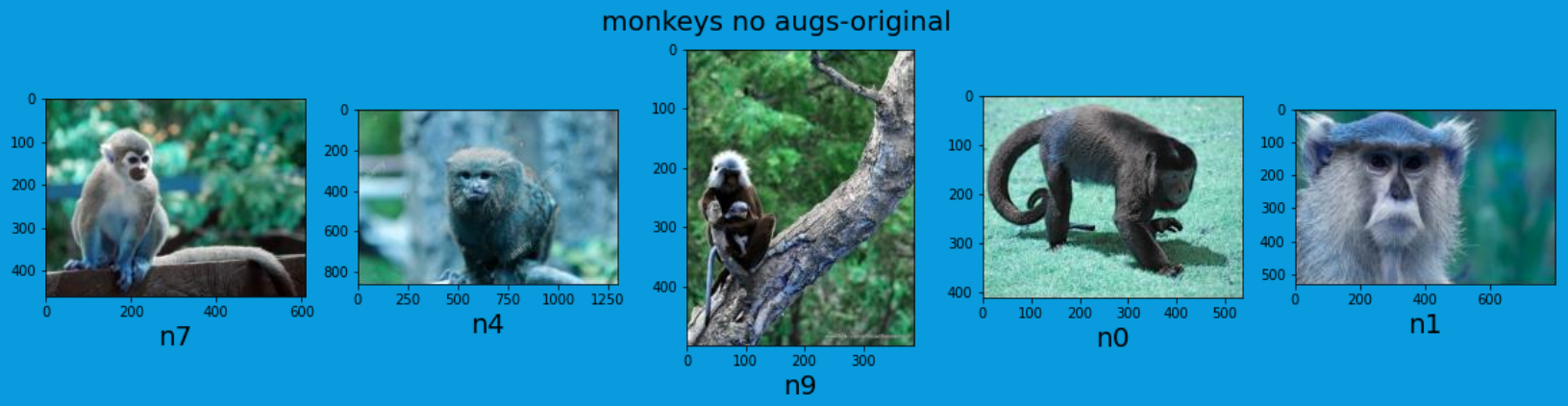




DATASETS - MONKEYS

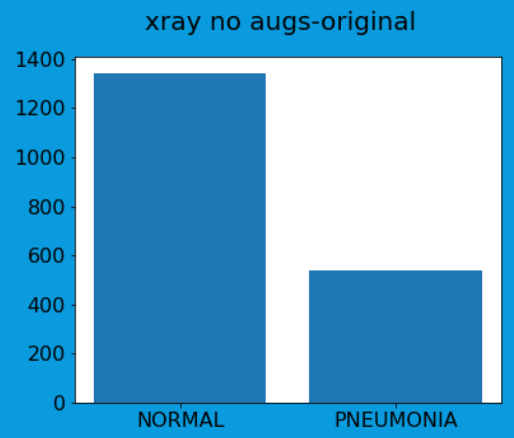
- 10 classes
- ~100 images of each class
- variable resolution

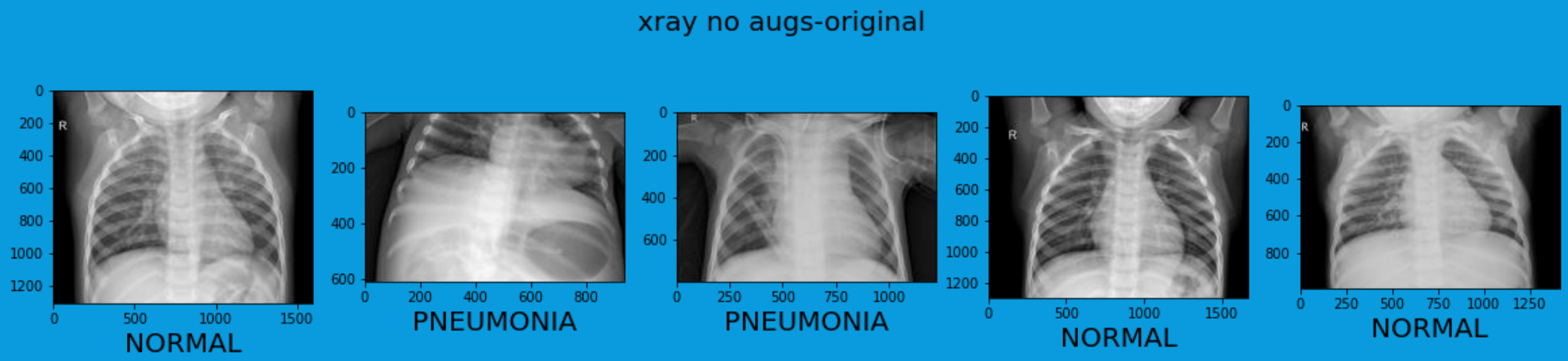




DATASETS - XRAY

- 2 classes
- Imbalanced
- variable resolution

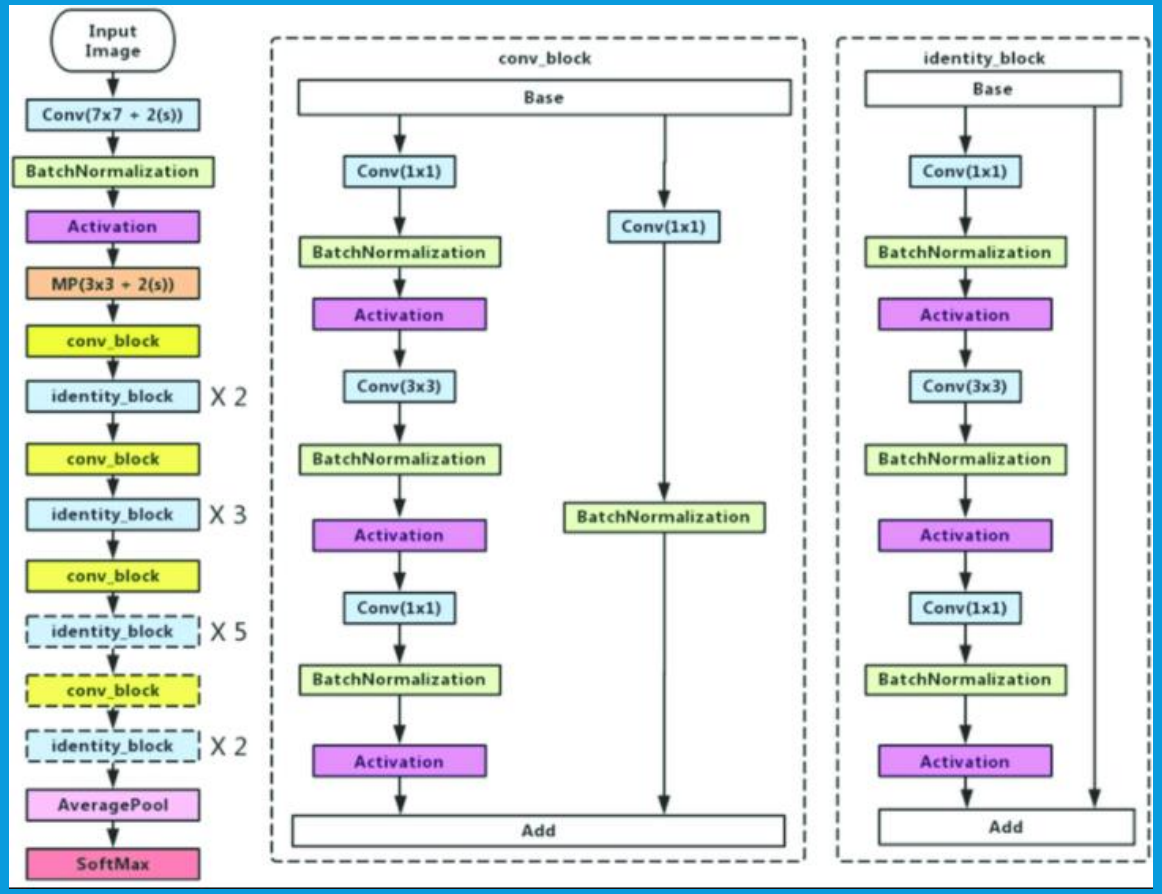




MODELS

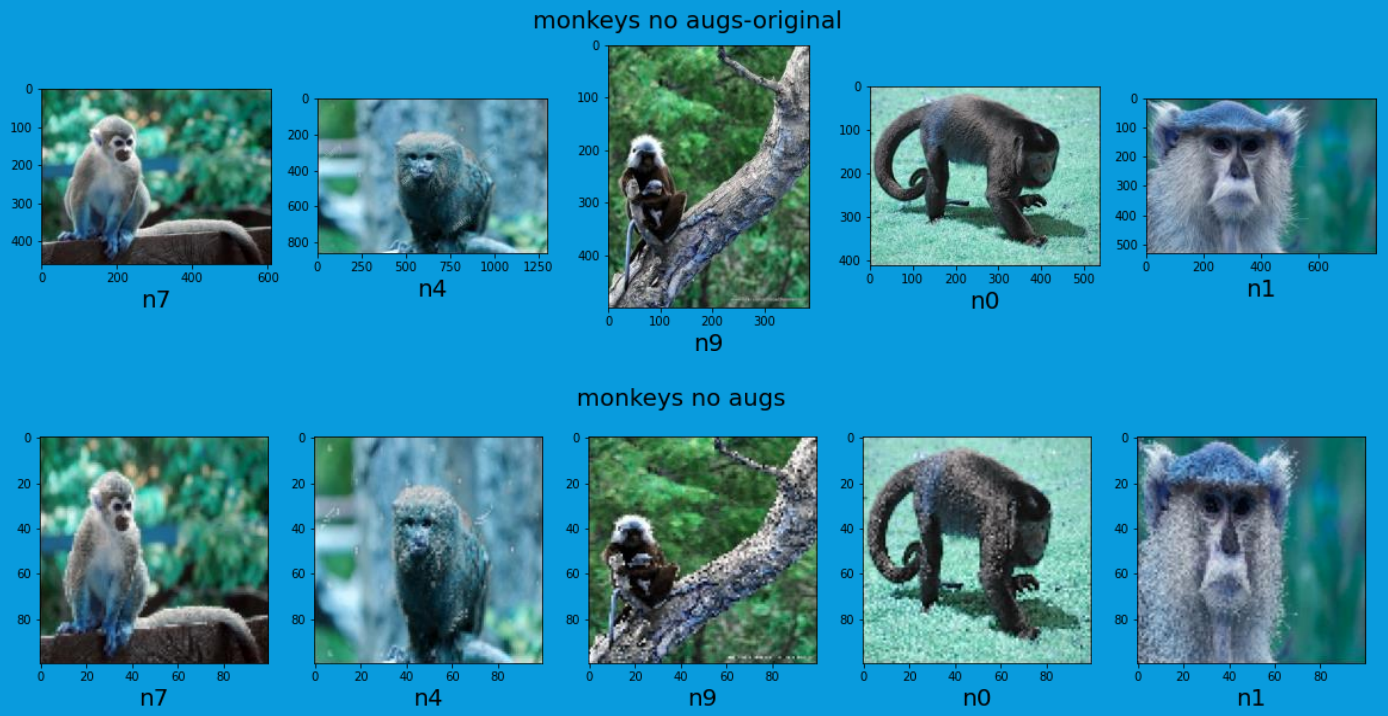
- Simple CNN
  - Conv -> Flatten -> Output
- Pooling CNN
  - Conv -> MaxPool -> Conv -> Maxpool -> Dropout -> Flatten -> Output
- Resnet50
  - Pretrained Resnet50 with imagenet dataset
  - Frozen layers -> Dense -> Dropout -> Dense -> Dropout -> Output

RESNET-50 ARCHITECTURE



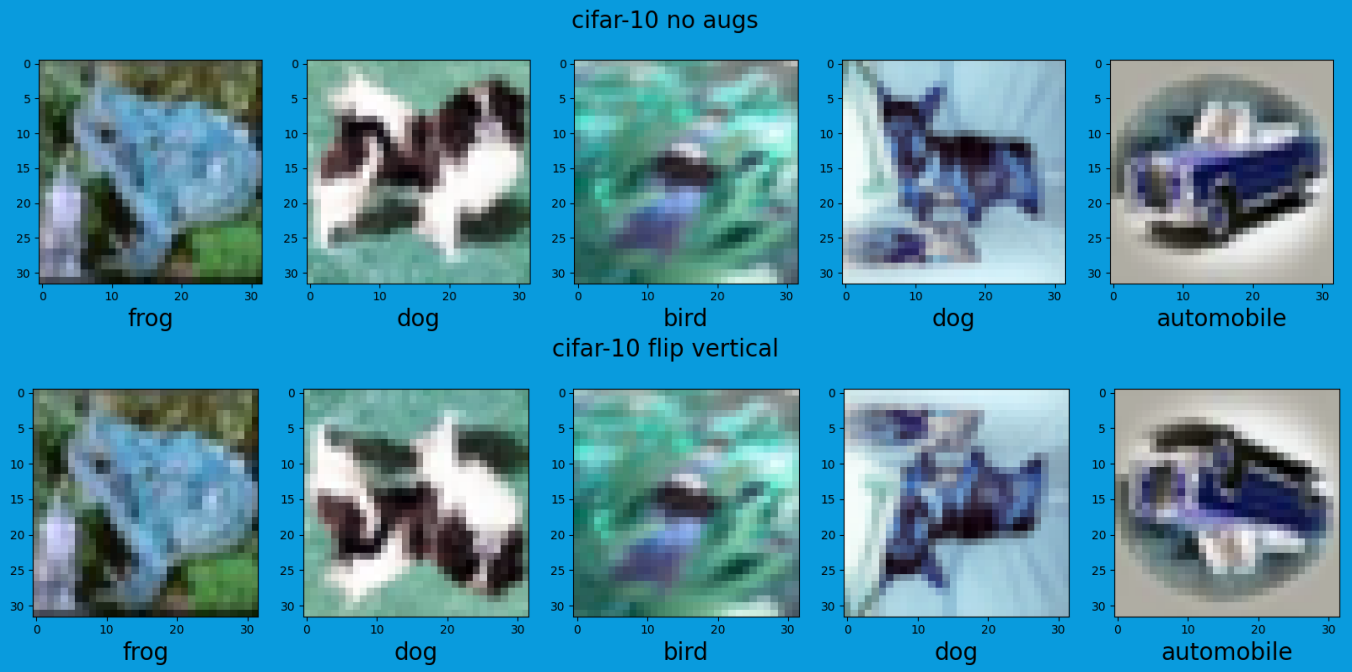
PREPROCESSING

- Preprocessing steps
  - Resize
    - Different resolutions
  - One-hot labels
  - Preprocessing
    - Normalize
    - Resnet preprocess



AUGMENTATIONS

- Experiments with 20 different augmentations, only 1 at a time
- What it means to augment an image?
  - Each image has 50% chance to be changed in training
  - No changes to test set

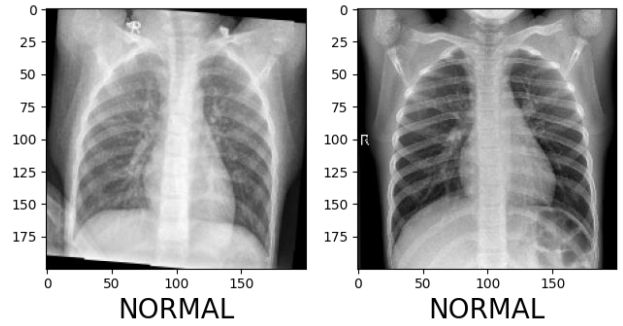


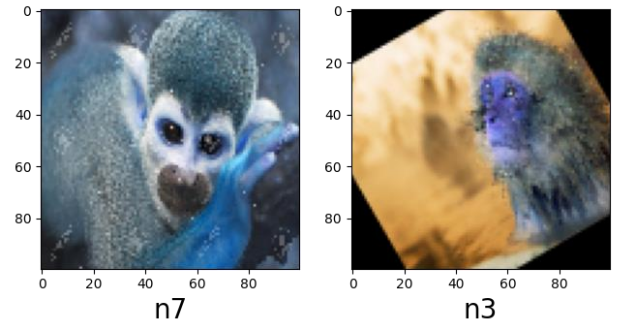
TOP 3 "BEST" AUGMENTATIONS

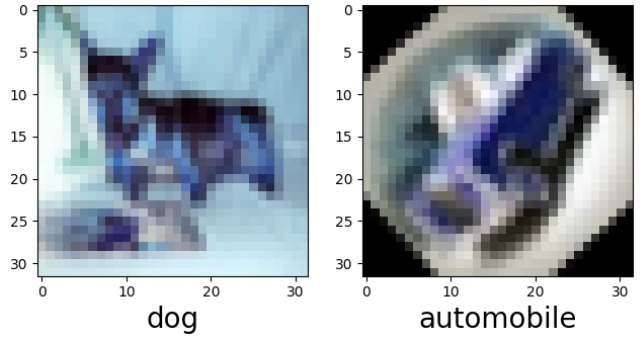
- Based on the absolute and relative increases in metrics
- Metrics:
  - Test accuracy
  - Test weighted precision
- 9 Cases for each augmentation
  - Every dataset-model combination

ROTATION

- Rotation of picture by –45 to 45 degrees
- Best for x-rays







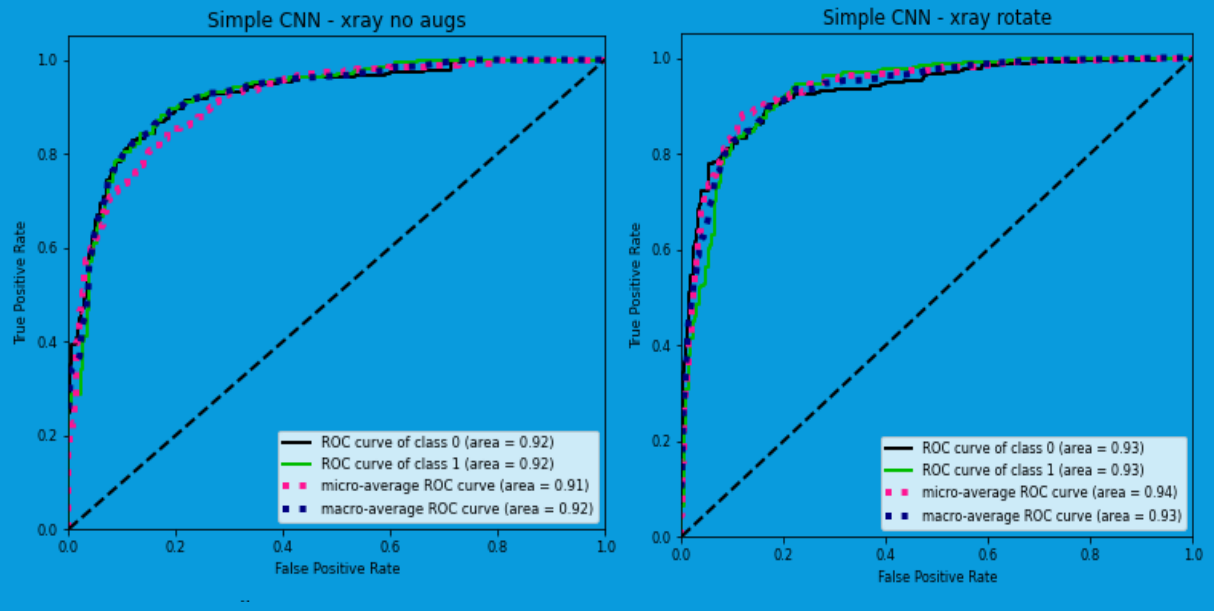
ROTATION XRAYS

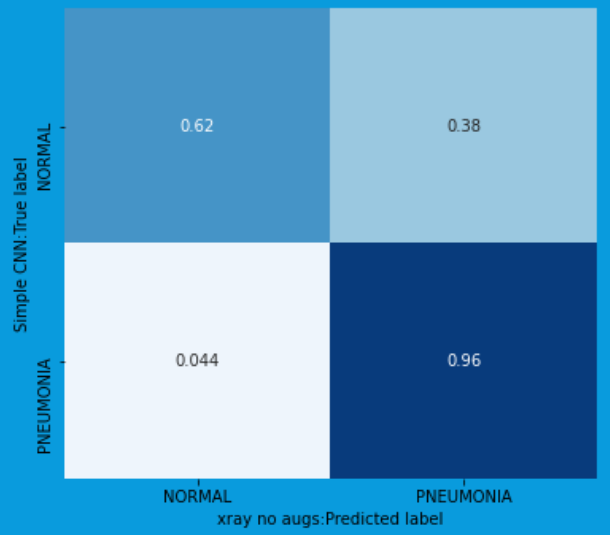
- Better than no-aug in 6/9 cases
- Best for simple and resnet model
- Biggest increase of all augmentations

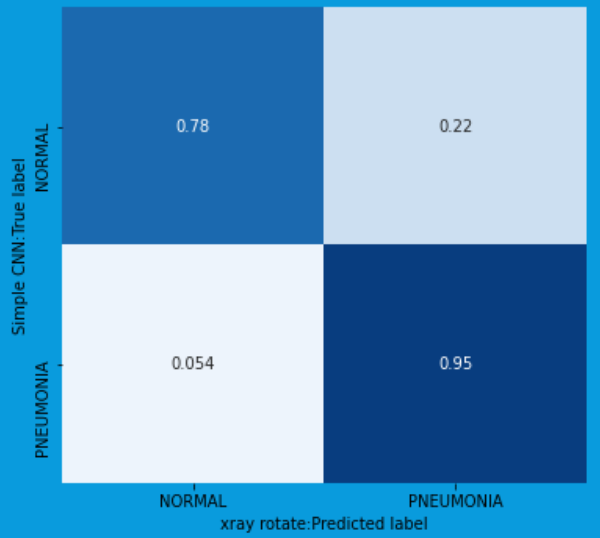
Simple model	No augs	Rotate
Accuracy	82.8	88.3
Weighted precision	83.9	88.4

Resnet model	No augs	Rotate
Accuracy	82	87.1
Weighted precision	84.6	88.3

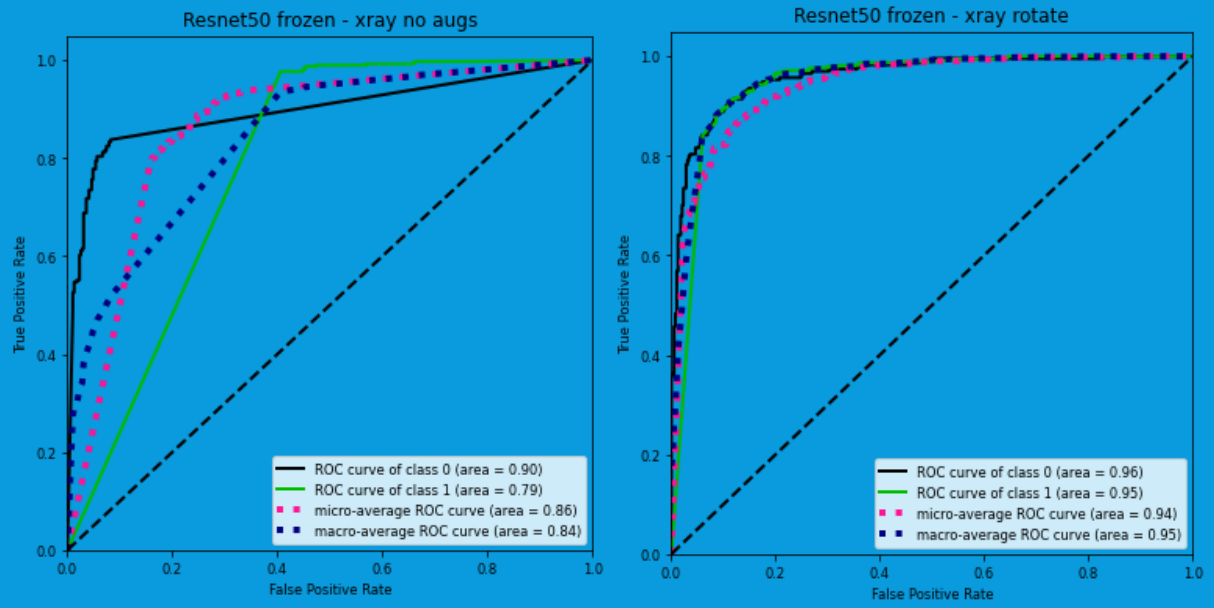
ROTATION FOR X-RAYS – SIMPLE CNN

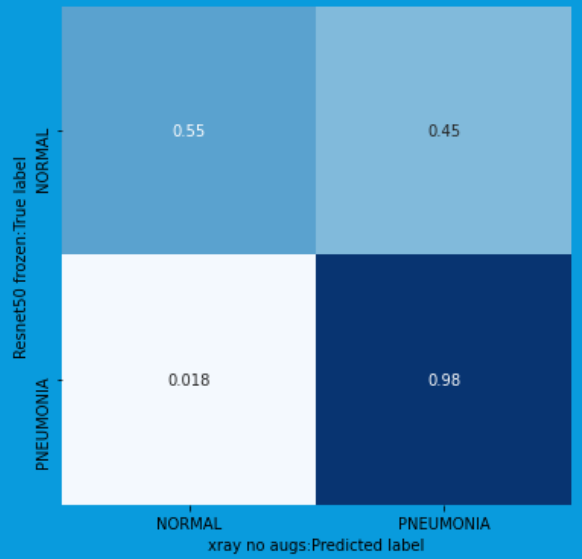


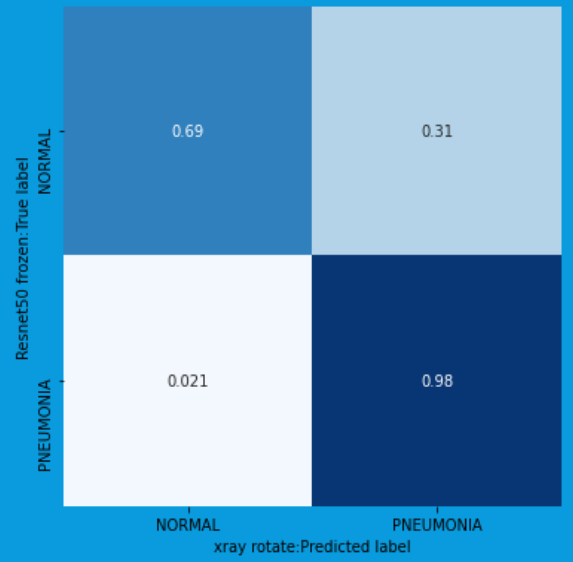




ROTATION FOR X-RAYS – RESNET50

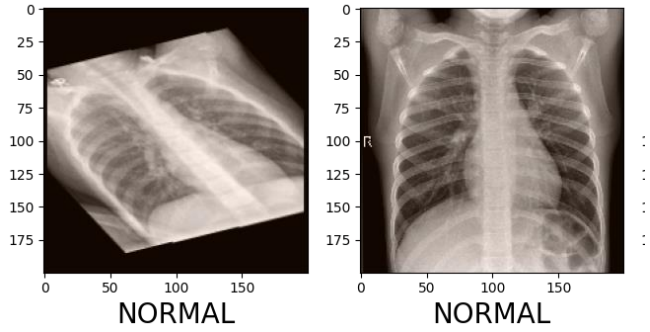


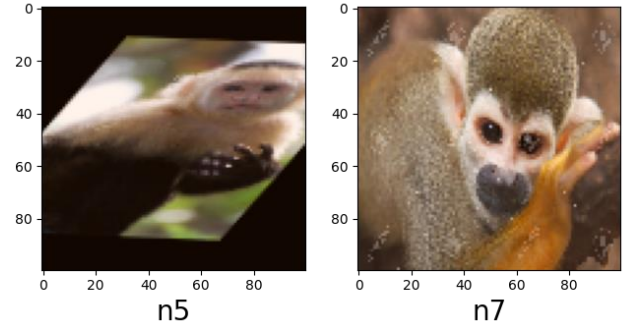


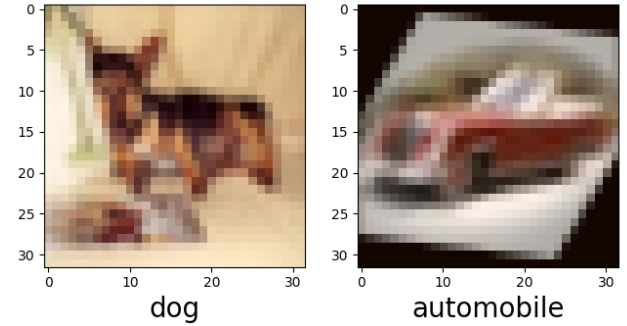


SHEARING

- Shearing of picture by –45 to 45 degrees
- Looks similar to rotation
- Again, best for xrays







SHEARING XRAYS

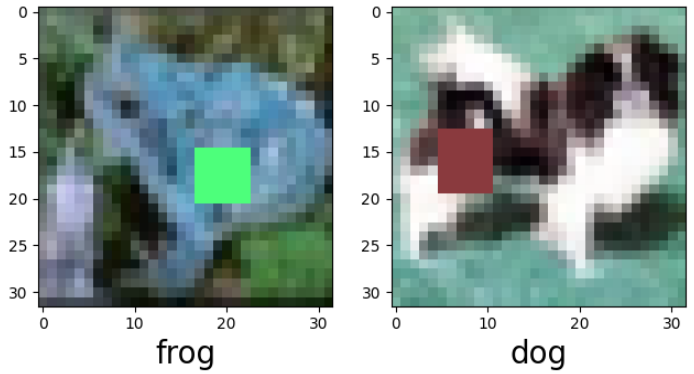
- (Again) Best for simple and resnet model
- Better than no-aug in only 4/9 cases
- Better results only for resnet model, but on each dataset (except simple model on xrays)
- In other models made accuracy worse by up to 4%

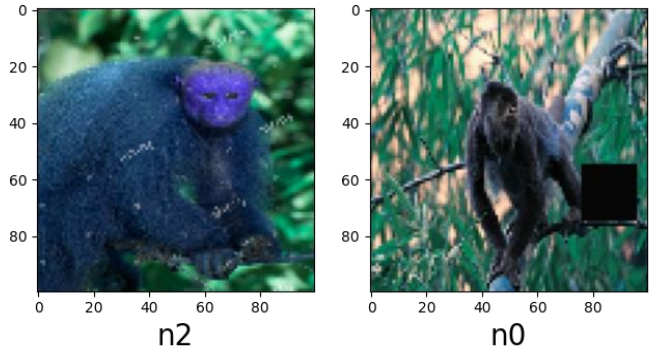
Simple model	No augs	Shearing (bracket is rotation acc)
Accuracy	82.8	87.5 (88.3)
Weighted precision	83.9	87.5 (88.4)

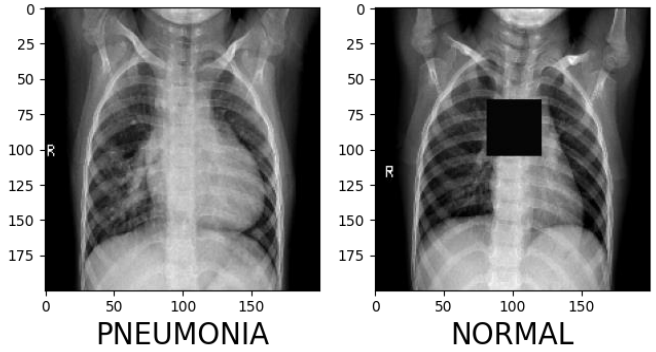
Resnet model	No augs	Shearing
Accuracy	82	86.3 (87.1)
Weighted precision	84.6	87.5 (88.3)

COLORED CUTOUT

Adding colored square into image
Best for xray and monkeys







COLORED CUTOUT

- Only beter on resnet
- Better than no-aug in only 3/9 cases
- Decreases accuracy on monkeys by 22% and 27 % on simple models
- Model matters

Resnet model Xrays	No augs	Colored cutout
Accuracy	82.8	86.2
Weighted precision	83.9	87.3

Resnet model Monkeys	No augs	Colored cutout
Accuracy	85.6	86.2
Weighted precision	86	89.1

FURTHER QUESTIONS

- Are augmentations good even on simpler models ?
- Best augmentations for specific dataset?
- Was no-aug best at something?
- Which augs were beneficial most of the time?

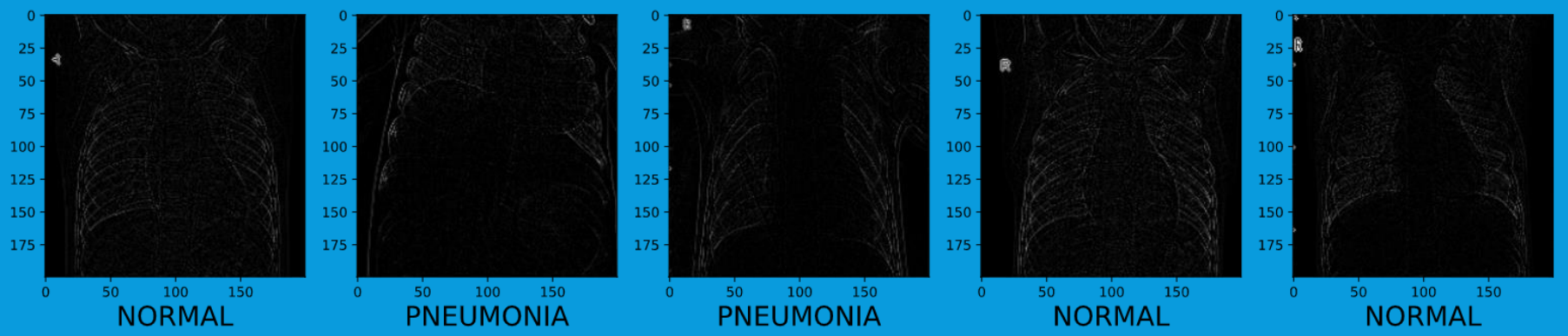
SIMPLE CNN

Dataset	Aug	Acc	Precision
Xray	Rotate	+5.5	+5.5
Xray	Shearing	+4.7	+3.6
Cifar-10	Horizontal flip	+4	+3.2
Xray	Edge map	+2.8	+1.8
Xray	Motion blur	+2.2	+1.4

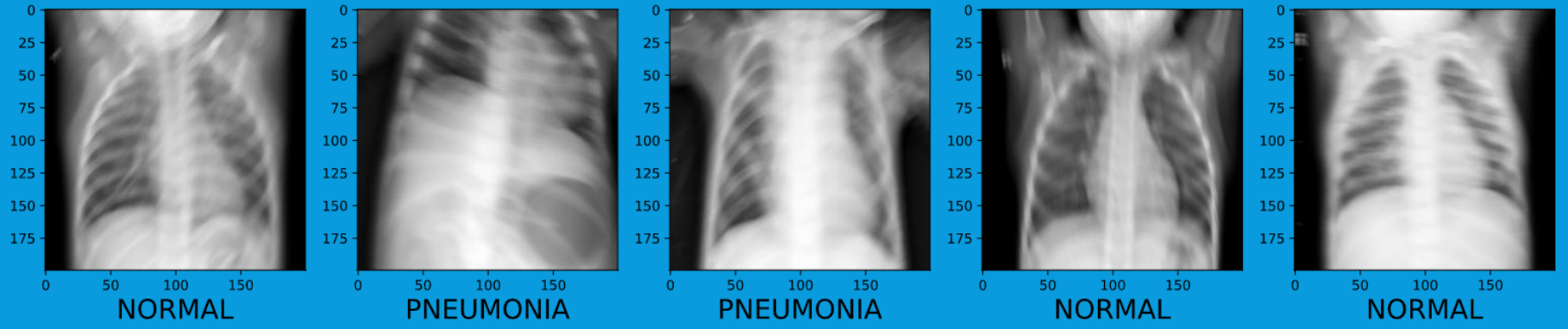
- Few percentages increases follow…
- Benefited by augmentations in 19/60 (32%) cases
- Benefited on x-ray and cifar-10 datasets

SIMPLE CNN

- Edge map



- Motion blur



POOLING CNN

- Only 4/20 augmentations had positive effect
- Is pooling just bad?
- 4/60 (6.6%) cases had positive impact
- Benefited on monkeys and xray

No-aug accuracy	Simple	Pooling	Resnet
Cifar - 10	60	71.4	66
Xray	82.8	82.5	82
Monkeys	61.7	66.5	85.6

Dataset	Aug	Acc	Precision
Monkeys	Rotate	+3.3	+3.5
Monkeys	Horizontal flip	+2.9	+2.7

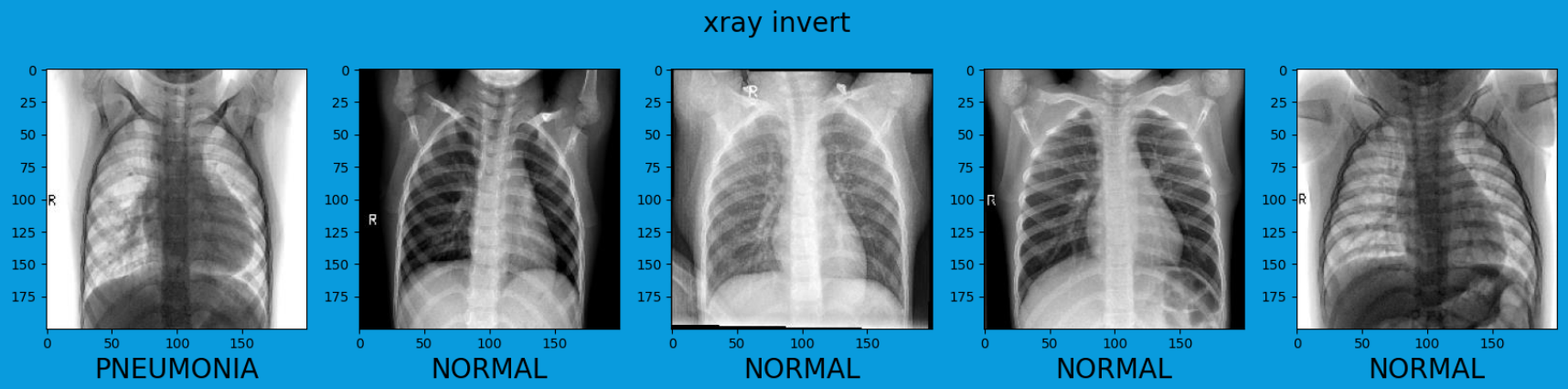
XRAY DATASET

- Most positively impacted by augmentations
- 31/60 (52%) cases had positive impact
- Every augmentation had at least 1 positive impact in some model
- Augs helped on simple and resnet model

	Acc	Precision
Rotate	+5.4	+4.5
Invert	+5.9	+4.3
Shearing	+4.6	+3.6

XRAY DATASET

- Invert – only good on Xrays



CIFAR-10 DATASET

- 26/60 (43%) had positive impact
- 17 augs had positive impact in at least one model
- Augs helped on simple and resnet model

	Acc	Precision
Horizontal flip	+4	+3.2
Rotate	+2	+0.5

- The rest under 1%
- Enough data, low resolution?

MONKEYS

- 11/60 (18%) cases had positive impact
- 9 augs had positive imapct on at least 1 model
- Augs helped on resnet and pooling models

	Acc	Precision
Rotate	+3.3	+3.5
Horizontal flip	+2.9	+2.7
Cutout	+2.9	+3

WHICH AUGS ARE MOST OFTEN USEFUL

- Ignores the actual increase, only looks at if helped/not helped
- 9 experiments for each aug

```
[('rotate', 6),
 ('flip horizontal', 5),
 ('translate random', 4),
 ('elastic', 4),
 ('shearing', 4),
 ('brightness', 4),
 ('cutout', 4),
 ('gaussian noise', 4),
 ('to edge map', 4),
 ('flip vertical', 3),
 ('gamma contrast', 3),
 ('clahe', 3),
 ('cutout colored', 3),
 ('pixel noise', 3),
 ('gaussian blur', 3),
 ('motion blur', 3),
 ('enhance edges', 3),
 ('channel adding', 2),
 ('dropout', 2),
 ('invert', 1),
 ('no augs', 0)]
```

WORST AUGMENTATIONS BY ACCURACY

Dataset	Aug	Acc
Monkeys	Clahe	-12
Monkeys	Gamma contrast	-11
Monkeys	Random translate	-10
Cifar-10	Vertical flip	-9
Xray	Horizontal flip	-7
Xray	Brightness	-7

THE END

- Choice of models, metrics, datasets, augmentations, resizing, training…
- Conclusions