CIFAR10 Noisy Image Classification

Contributors: Christie Du, Fucheng Liu, Vaishak Naik, Yiming Xia, Shanyue Zeng

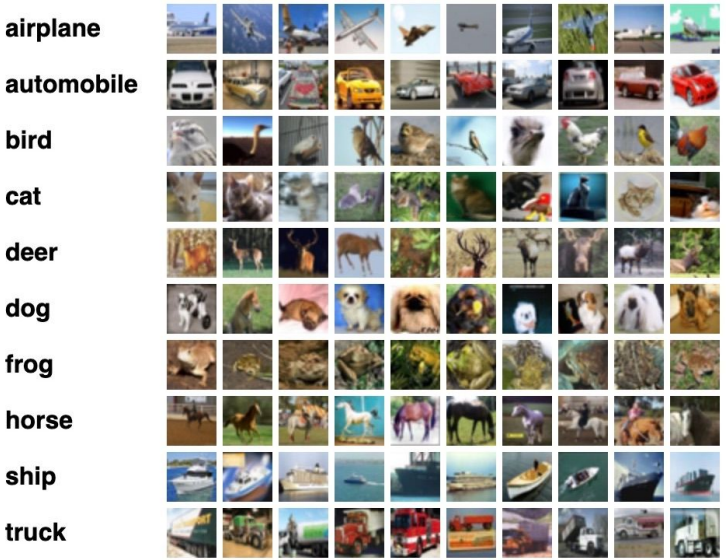
Problem

Image classification with the **noisy version** of "CIFAR-10" dataset

- Design a sophisticated classification model
- Address the label noise issue

CIFAR-10 Dataset

- Noisy labels for 40,000 images
- Clean labels for 10,000 images



Data Augmentation

Five types of data augmentation on 10K clean data → 50K augmented data:

- flip (hflip or vflip), rotate (90˚, 180˚, 270˚), distort, lighting (brighter, darker), color

Original | Flip | Rotate | Distort | Lighting | Color

Total processed data: 100K images (50K original, 50K augmented). 2.46GB (npy files)

Baseline Performance: Logistic Regression

The baseline LR model is only trained on the 50K uncleaned data:

- Test Accuracy: 24%
- Avg Precision: 24%
- Avg Recall: 24%

The accuracy is a little better than randomly predicted (10%)

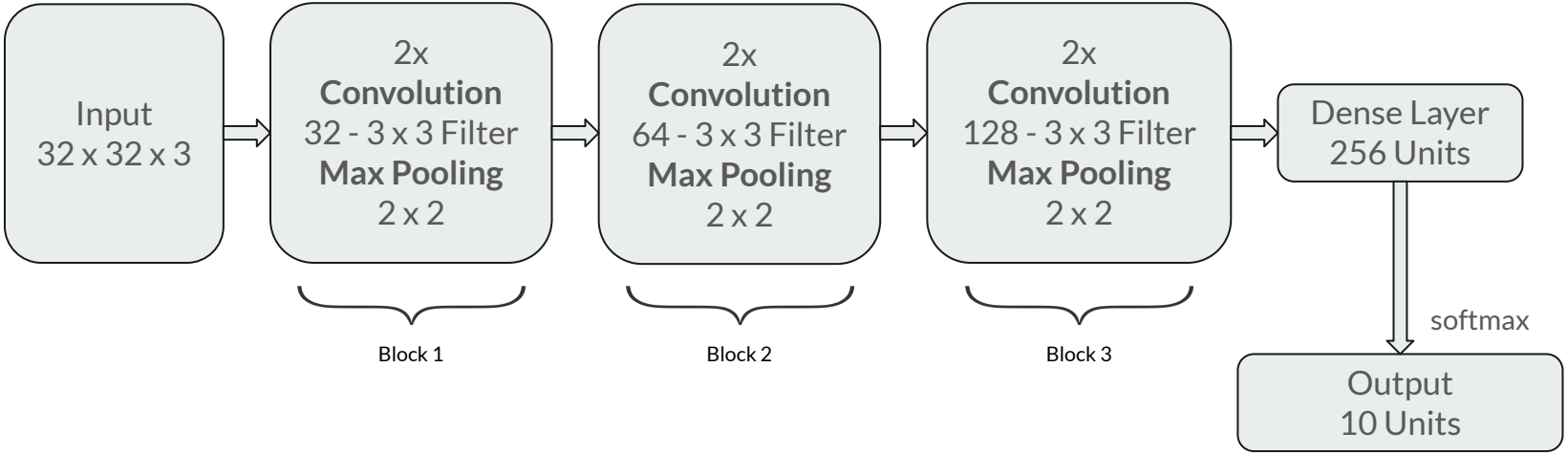
LR model (75K train, 25K validation, random split)

- Test Accuracy: 20%
- Avg Precision: 20%
- Avg Recall: 20%

Predictive Models: Model I & Model II

Model I : Architecture

CNN



Model I: Performance

Model Accuracy/Loss

	Accuracy	Loss
Train	0.6731	1.0710
Validation	**0.6051**	1.4852

Runtime Metrics

Computing Type	Training Time
Apple M1 CPU	~34 min
Tesla K80 GPU	~9 min
Intel(R) Xeon(R) CPU @ 2.20GHz	~225 min / ~ 3.75 hr

Model II: Workflow

CNN+ semi-supervised learning

Train model to clean noisy labels	Predict labels for noisy data	Train model II on cleaned data
The same architecture for model I is trained on clean + augmented data only (60K images). We call this model `clean_model`.	`clean_model` is used to predict new labels for noisy data (40K images).	The same architecture for model I is trained using newly cleaned labels.

Candidate Model II: Transfer learning, pre-trained VGG16, training on one additional layer

Model II: Performance

Model Accuracy/Loss

Clean_label model	Accuracy	Loss
Train	0.9240	0.2233

Model II	Accuracy	Loss
Train	0.9031	0.2680
Validation	**0.8266**	0.5544

Runtime Metrics

Model	CPU Apple M1 Training Time	GPU Tesla K80 Training Time	GPU Intel(R) Xeon(R) CPU
Clean model	~ 26 min	~ 4 min	~ 3 hr
Label model	~ 10 min		
Model II	~ 34 min	~ 9 min	~ 4 hr

Conclusion

Model	Validation Accuracy	Total Runtime (i9, Apple M1, Tesla K80)	Predict 25K Runtime (i9, Apple M1, Tesla K80)
Baseline	20%	~10 s	~2 s
Model I	60.51%	(~3.75 hr, ~35 min, ~9 min)	(~12 min, ~3 min, ~3s)
Model II	82.66%	(~8 hr, ~1 hr, ~13 min)	(~12 min, ~3 min, ~3s)

Pros

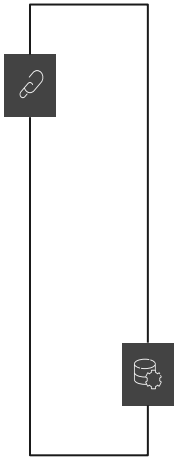
Higher accuracy

Higher precision

Higher recall

Fast on prediction

Only need to train once



Cons

Augmented data doubles data size

Significantly slower train time

Questions?