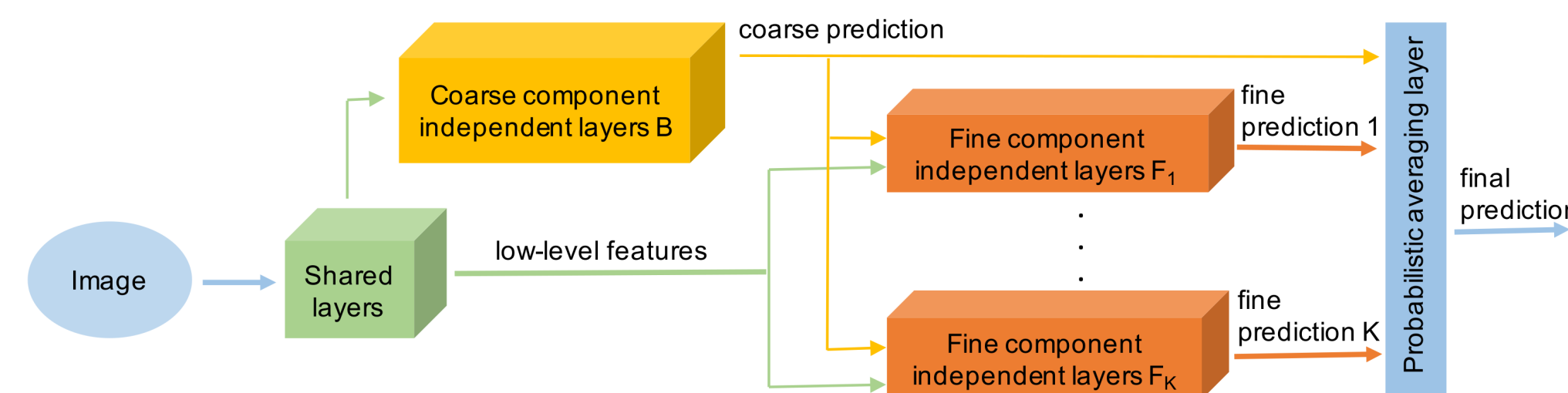


Hierarchical Image Classification

Prashant Kumar
Stanford University

Introduction

Can image classification via CNN benefit from Hierarchical classification?



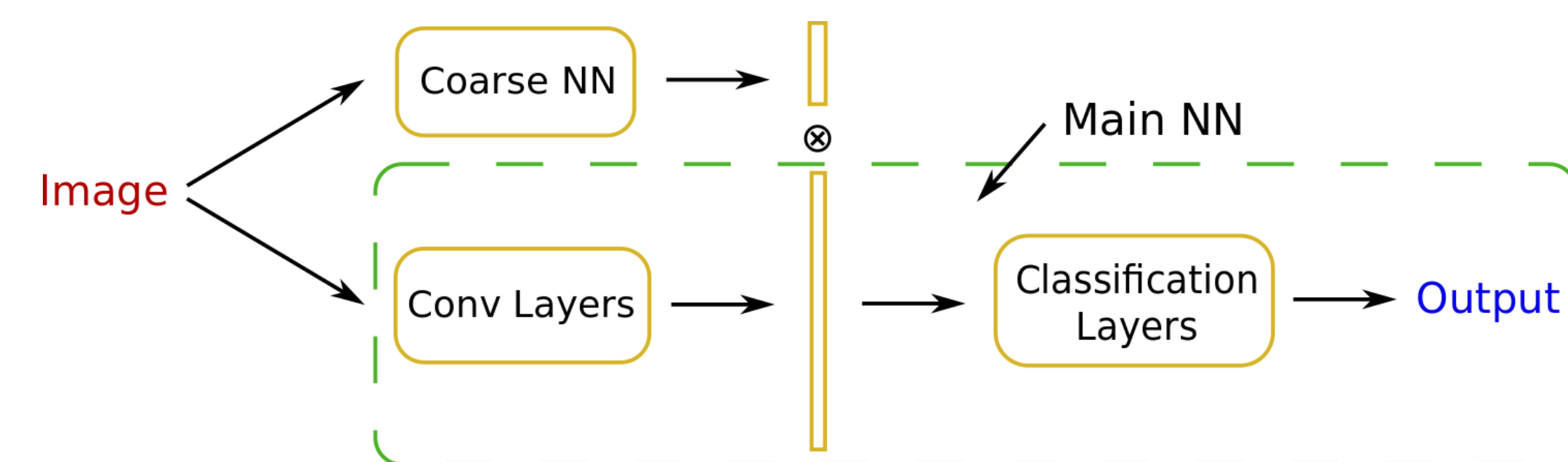
HD-CNN

Independent layers for fine-classification \implies Depends on coarse classification accuracy.

Bilinear-CNN utilizes correlations between filters of convolutional network.

Course project idea: Use correlations between filters and coarse categorization.

CNN architecture



Coarse NN: VGG13

Main NN: VGG16

Methods:

1. Category Hierarchy (CH)
2. Coarse Visual Category (CVC)
3. End-to-End (ETE)

Hyperparameters:

Trainset:Devset per label	80:20
Number of coarse classes	10
Initial Learning rate	0.01
L2 regularization	0.001

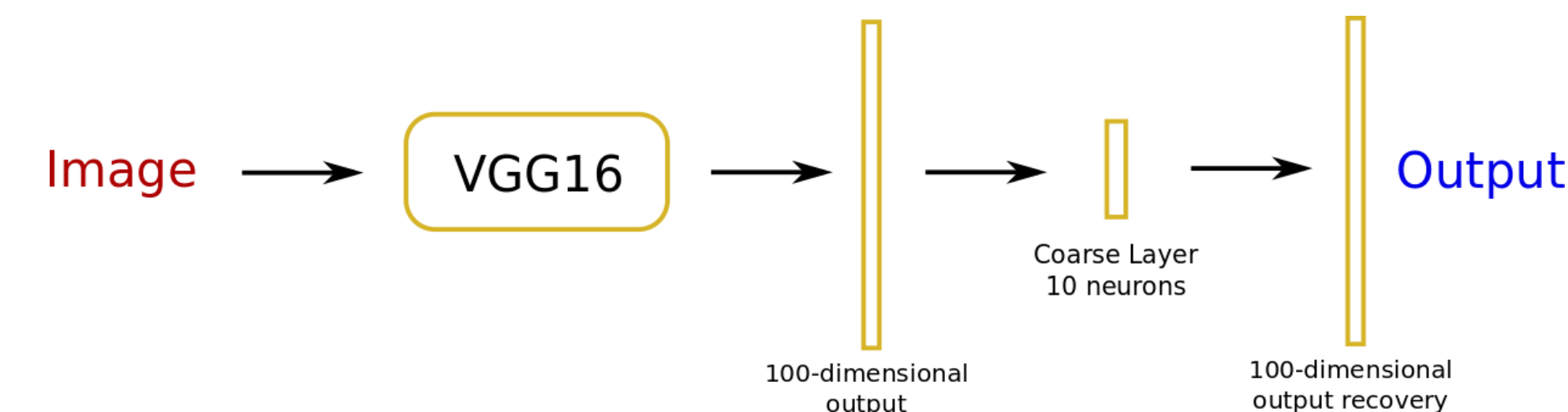
VGG16 baseline

VGG16 baseline on CIFAR100

Dataset	Accuracy
Trainset	94.37%
Validation set	69.03%
Test set	68.74%

31 epochs.

Category Hierarchy



Category Hierarchy	
Trees	chair, forest, maple tree, motorcycle, oak tree, orange, pine tree, snake, willow tree, worm
Vehicles	bus, cattle, fox, lawn mower, mountain, palm tree, pickup truck, streetcar, tank, tractor, train
Big Carnivores	apple, leopard, lion, tiger, wolf
Category 4	bowl, can, clock, keyboard, lamp, plate, raccoon, rocket, skunk
Humans	baby, bicycle, boy, cloud, flatfish, girl, man, plain, possum, sea, woman
Sea animals	crab, dinosaur, dolphin, lobster, otter, ray, seal, shark, trout, turtle, whale
Category 7	bed, road, table, television, wardrobe
Category 8	aquarium fish, bear, beaver, camel, chimpanzee, elephant, hamster, rose
Insects	bee, beetle, bottle, butterfly, caterpillar, cockroach, couch, cup, orchid, pear, poppy, spider, sunflower, sweet pepper, telephone, tulip
Category 10	bridge, castle, crocodile, house, kangaroo, lizard, mouse, mushroom, porcupine, rabbit, shrew, skyscraper, snail, squirrel

Training Results

Coarse NN accuracies	
Dataset	Accuracy
Trainset	82.09%
Validation set	75.90%
Test set	75.88%

31 epochs.

Overall accuracies	
Dataset	Accuracy
Trainset	84.46%
Validation set	69.90%
Test set	65.45%

66 epochs.

Coarse Visual Categories

Coarse NN accuracies

Dataset	Accuracy
Trainset	81.92%
Validation set	77.72%
Test set	77.61%

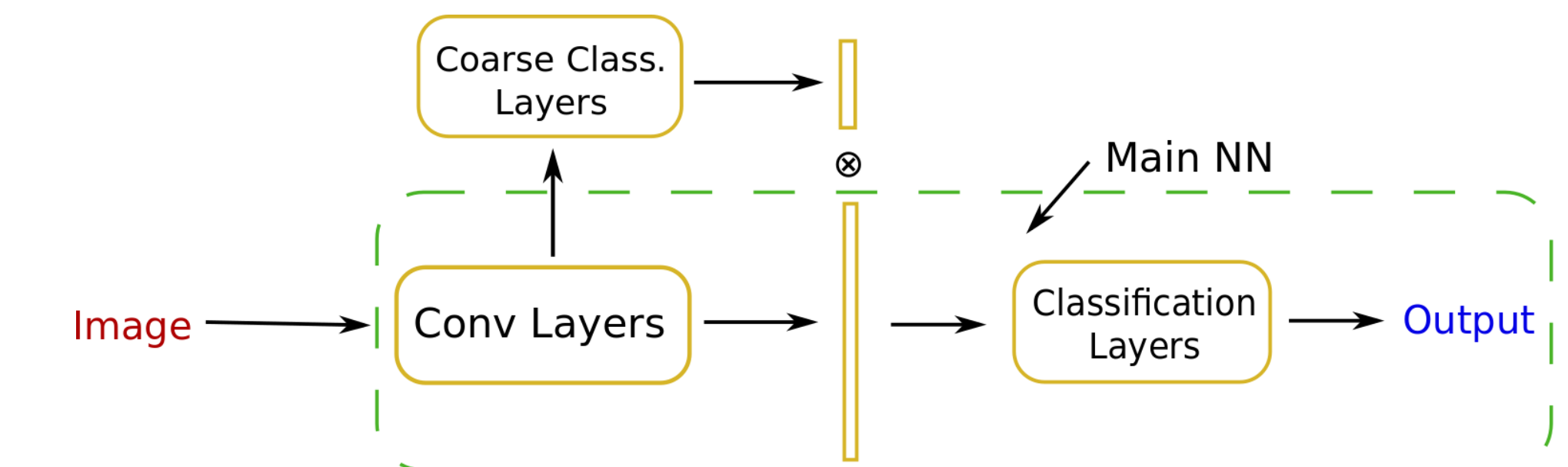
32 epochs.

Overall accuracies

Dataset	Accuracy
Trainset	85.01%
Validation set	67.22%
Test set	65.71%

76 epochs.

End-to-End



Overall accuracies

Dataset	Accuracy
Trainset	92.56%
Validation set	68.94%
Test set	68.42%

Conclusions

- Overfitting is a serious issue in Coarse NN.
- No significant change in accuracies.

References

1. Tsung-Yu Lin and Aruni Roy Chowdhury and Subhransu Maji. Bilinear CNN Models for Fine-grained Visual Recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1449–1457, 2015.
2. Z. Yan, H. Zhang, R. Piramuthu, V. Jagadeesh, D. DeCoste, W. Di, and Y. Yu. HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 2740–2748, 2015.
3. <https://github.com/kuangliu/pytorch-cifar>
4. <https://gist.github.com/jeasinema/ed9236ce743c8efaf30fa2ff732749f5>