Improved Image Classification with Coarse and Fine Labels

Anuvabh Dutt, Denis Pellerin, Georges Quénot

Motivation

Motivation

- Get meaningful probabilities for multimedia IR tasks
  - AVS, MED 0Ex

Motivation

- Get meaningful probabilities for multimedia IR tasks
  - AVS, MED 0Ex
- Convert query into a list of target concepts with weights and possibly Boolean formula.

Motivation

- Get meaningful probabilities for multimedia IR tasks
    - AVS, MED 0Ex
- Convert query into a list of target concepts with weights and possibly Boolean formula.
- Need to calibrate probabilities for using (and combining) concepts from various models
    - ImageNet, Places, etc

Motivation

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- **Classification**: Rank concepts given an image.
  - Metric: Top-K error. Normalization: soft-max.

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- Classificatior an image.
    - ✴ Metric: Top soft-max.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- Classification: Rank concepts given an image.
  - ✳ Metric: Top-K error. Normalization: soft-max.

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- **Classification**: Rank concepts given an image.
  - ✴ Metric: Top-K error. Normalization: soft-max.
- **Retrieval**: Rank images given a concept.
  - ✴ Metric: MAP. Normalization: Platt.

Motivation

- Scores ranges from $-\infty$ to $+\infty$
- Probabilities are expected to range from 0 to 1
- Sigmoid transform: $p(score) = 1/(1+e^{(A*score+B)})$
- Additional hint: among the samples within a small interval around p, a fraction of about p would have positive labels
- Platt's method: learn A and B by cross-validation to optimally satisfy the above hint

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- **Classification**: Rank concepts given an image.
  - ✴ Metric: Top-K error. Normalization: soft-max.
- **Retrieval**: Rank images given a concept.
  - ✴ Metric: MAP. Normalization: Platt.

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- **Classification**: Rank concepts given an image.
  - ✴ Metric: Top-K error. Normalization: soft-max.
- **Retrieval**: Rank images given a concept.
  - ✴ Metric: MAP. Normalization: Platt.
- In both cases: average over a number of images to be categorized or over a number of concepts to be retrieved.

Motivation

- Utilise semantic relationships among concepts for tasks like classification and retrieval.
- Classification: Rank concepts given an image.
  - ✴ Metric: Top-K error. Normalization: soft-max.
- Retrieval: Rank images given a concept.
  - ✴ Metric: MAP. Normalization: Platt.
- In both cases: average over a number of images to be categorized or over a number of concepts to be retrieved.
- Refine concept categories, from coarse to fine grained.

Related Work

Related Work

- YOLO9000: Better, Faster, Stronger
  Joseph Redmon, Ali Farhadi [arxiv:1612.08242]

Related Work

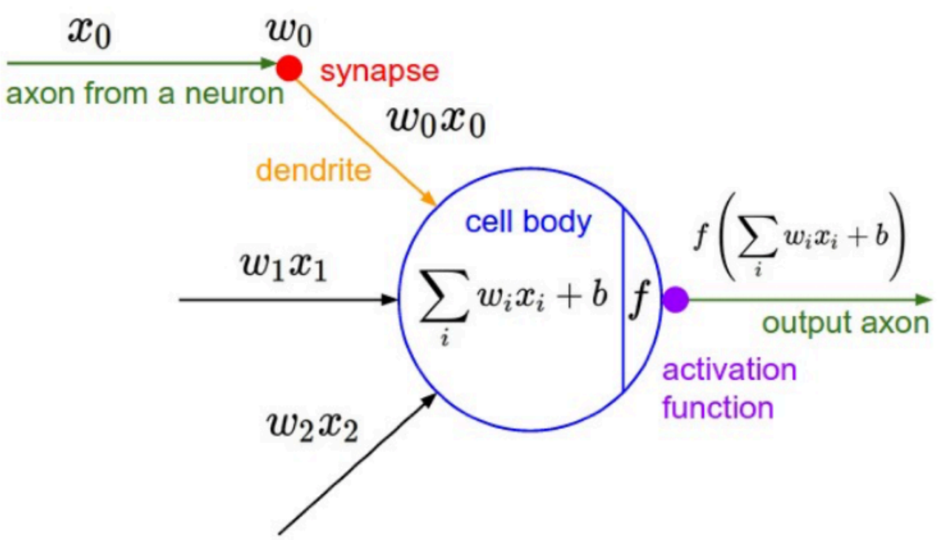
- YOLO9000: Better, Faster, Stronger
  Joseph Redmon, Ali Farhadi [arxiv:1612.08242]

- Classifier adaptation at prediction time.
  Amelie Royer and Christoph H Lampert. CVPR, 2015.

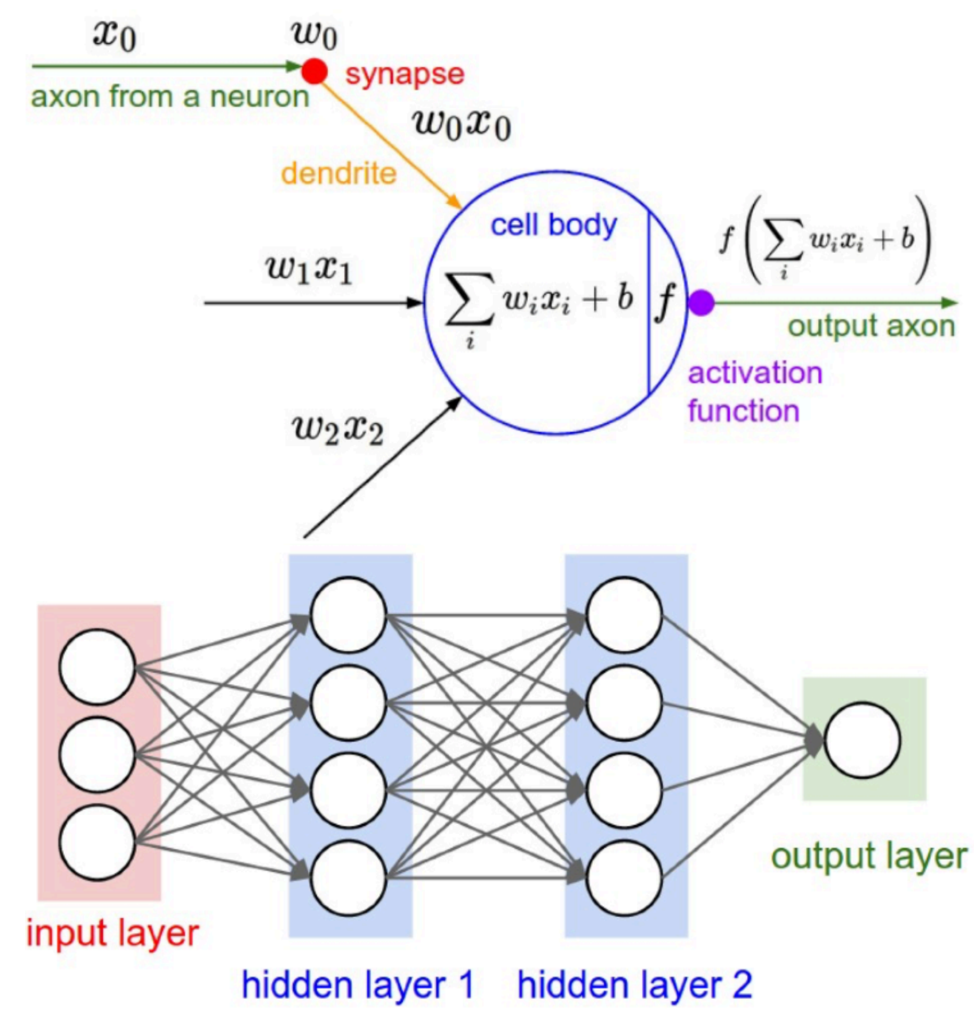
Background

Background

- Neural Network



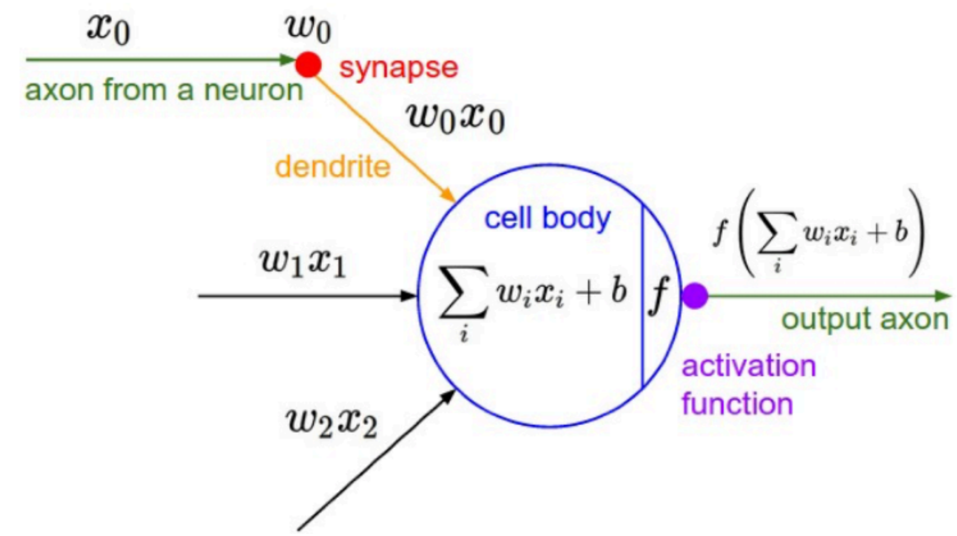
Background

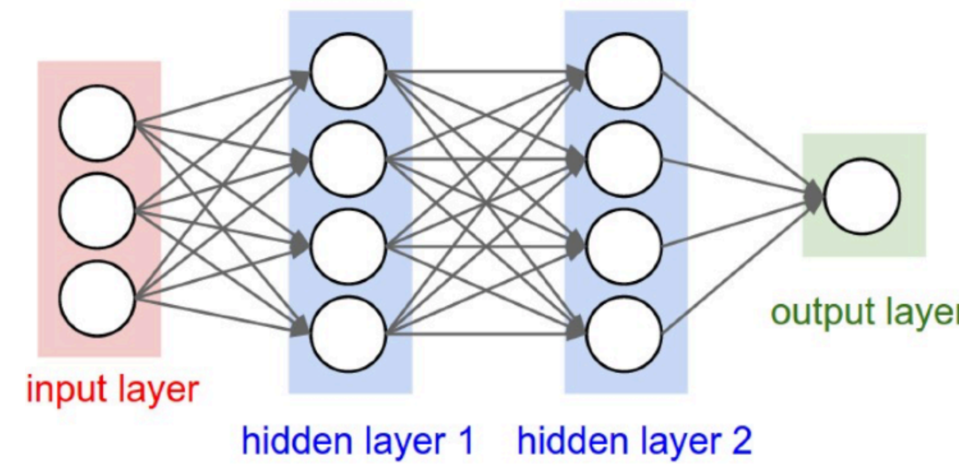


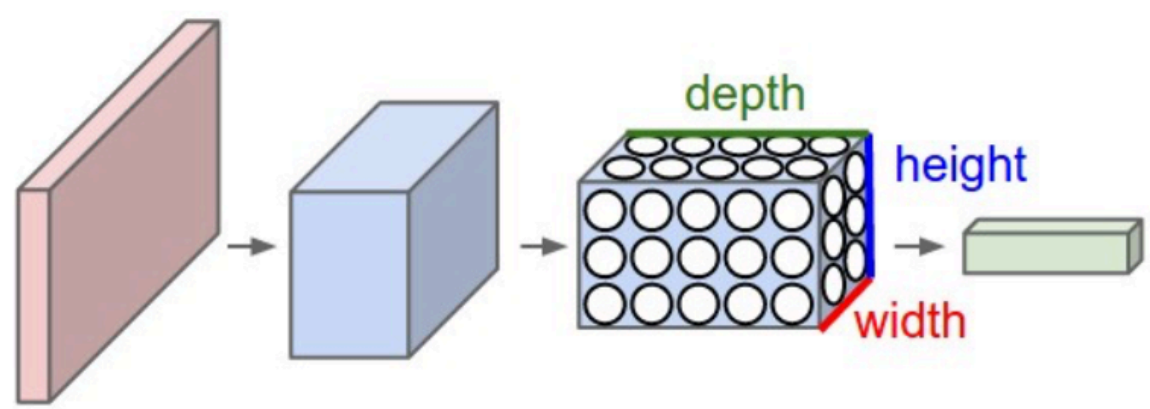
- Neural Network
- Deep Neural Network

Background

- Neural Network
- Deep Neural Network
- Deep Convolutional Network







Background

Background

- Train using stochastic gradient descent

Background

- Train using stochastic gradient descent
  - ✳ Gradients obtained from loss on mini-batches

Background

- Train using stochastic gradient descent
  - ✳ Gradients obtained from loss on mini-batches
- Regularization using `dropout`, `batchnorm`.

Background

- Train using stochastic gradient descent
  - ✴ Gradients obtained from loss on mini-batches
- Regularization using `dropout`, `batchnorm`.
- State-of-the art models in vision: ResNet, DenseNet

Background

- Train using stochastic gradient descent
  - ✴ Gradients obtained from loss on mini-batches
- Regularization using `dropout`, `batchnorm`.
- State-of-the art models in vision: ResNet, DenseNet
- After training, the classifier predicts a probability distribution over the categories.

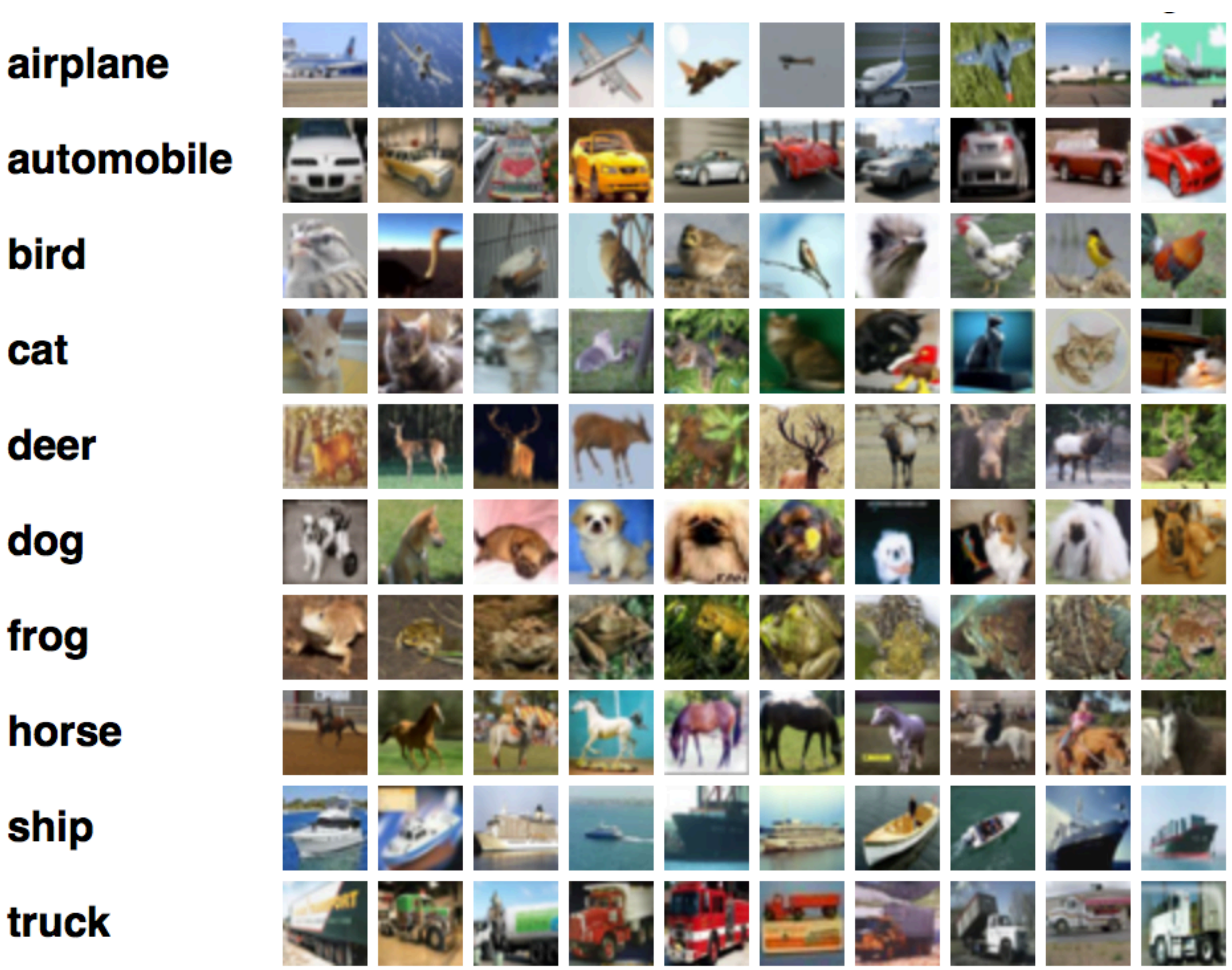
Background

- Train using stochastic gradient descent
  - ✴ Gradients obtained from loss on mini-batches
- Regularization using `dropout`, `batchnorm`.
- State-of-the art models in vision: ResNet, DenseNet
- After training, the classifier predicts a probability distribution over the categories.
  - ✴ The value assigned to a category is the “confidence” of the classifier for that category

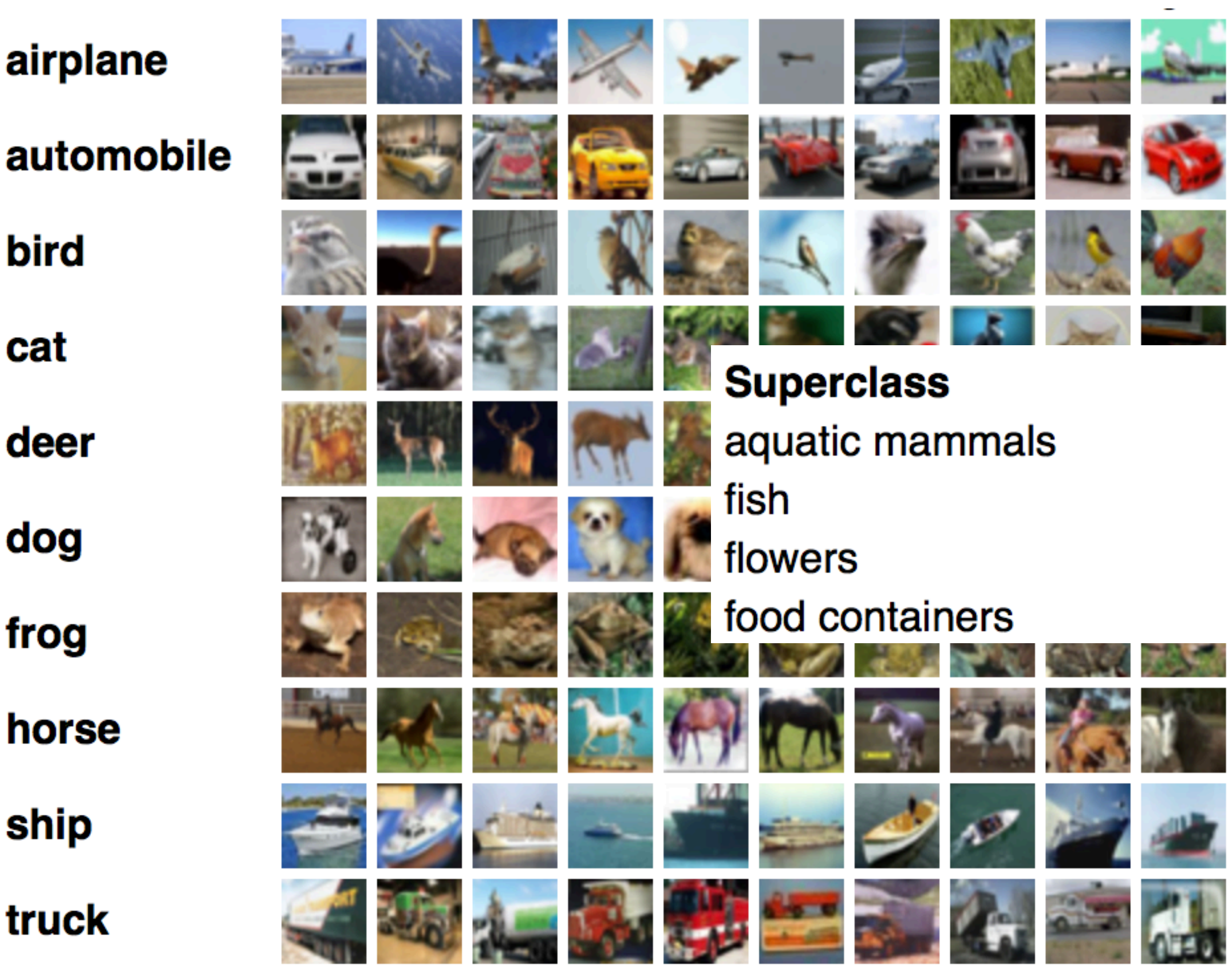
Problem Setting

- Image classification
- Labels with semantic relationship
- Example:
  - Coarse label: vehicle
  - Corresponding fine label(s): car, truck, bike

Problem Setting



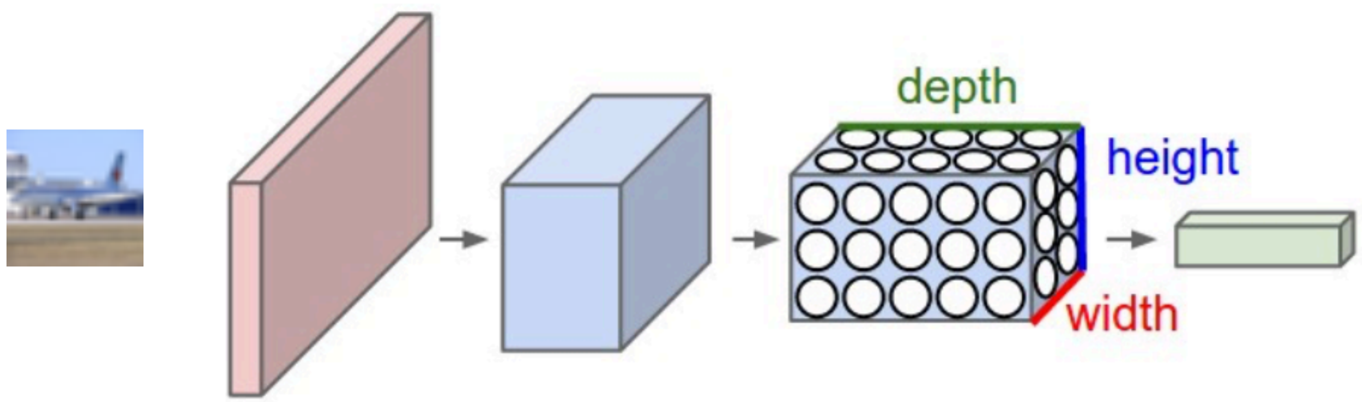
Problem Setting



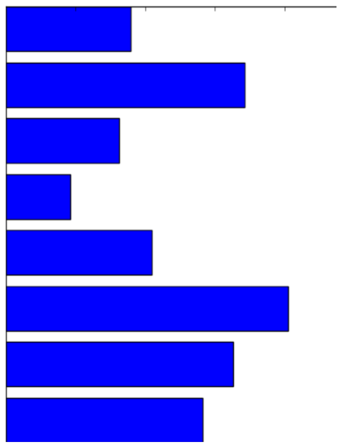
Superclass	**Classes**
aquatic mammals	beaver, dolphin, otter, seal, whale
fish	aquarium fish, flatfish, ray, shark, trout
flowers	orchids, poppies, roses, sunflowers, tulips
food containers	bottles, bowls, cans, cups, plates

Classifier Confidence

Classifier Confidence

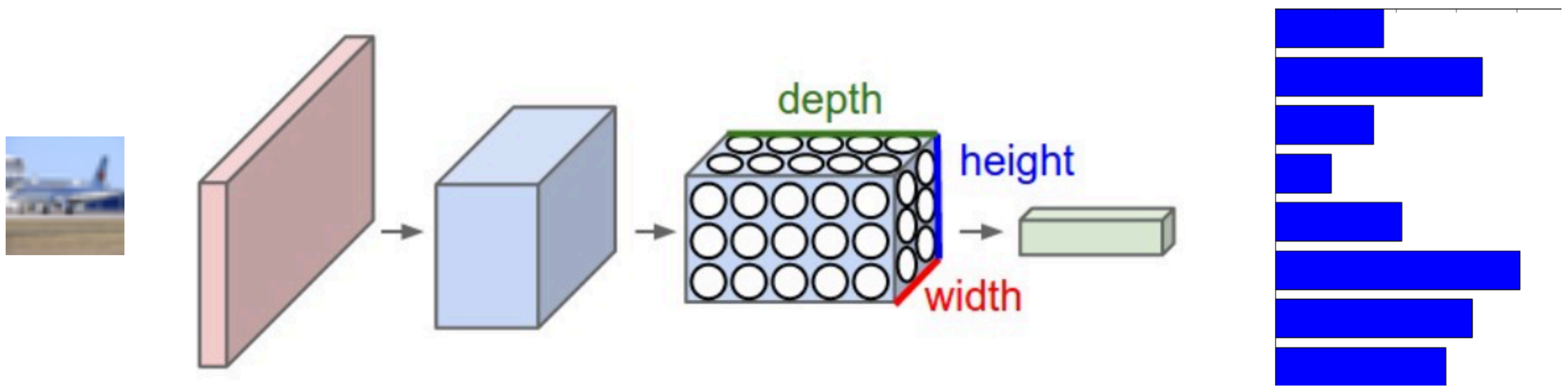


Fine

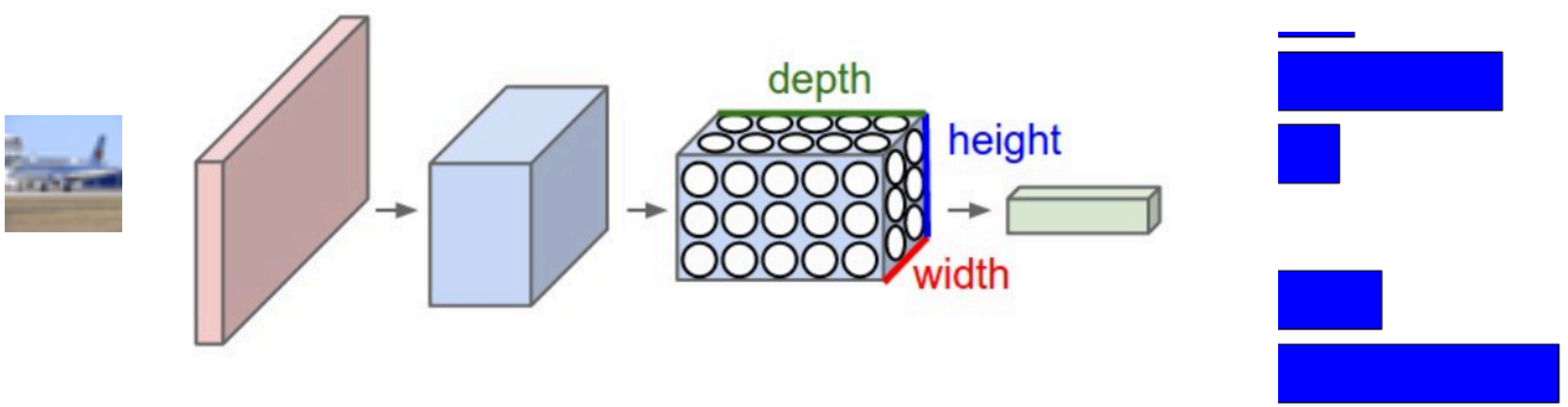


Classifier Confidence

Fine



Coarse



Motivation

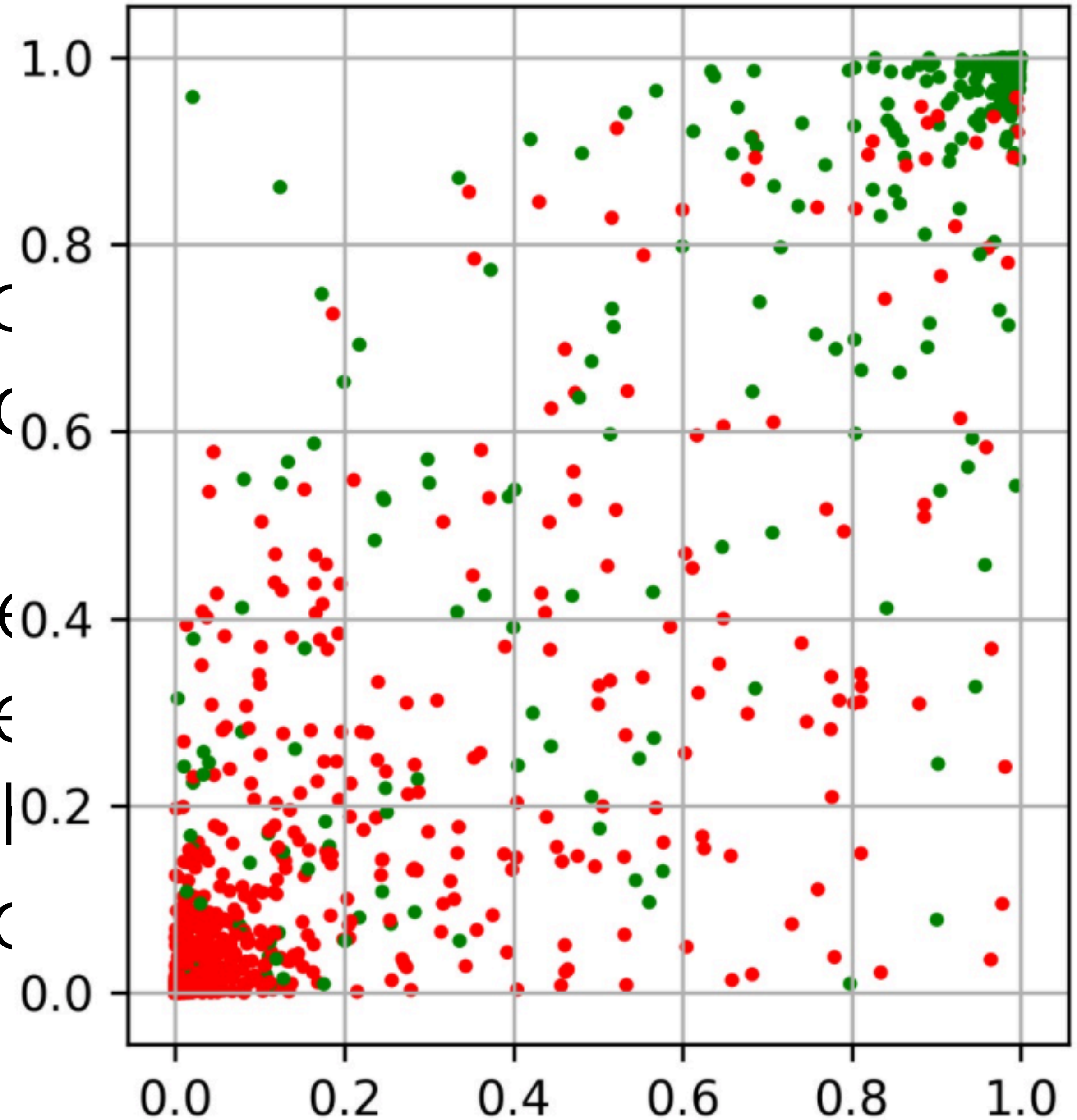
- Trained models have a confidence value for the coarse and fine categories, respectively.

- If the model captures the ***semantic*** structure, the confidence in the coarse category should be similar to the total confidence in the corresponding fine categories.

Motivation

- Trained mc ıe for the coarse anc y.

- If the mode ıcture, the confidence ıld be similar to tl rresponding fine catego

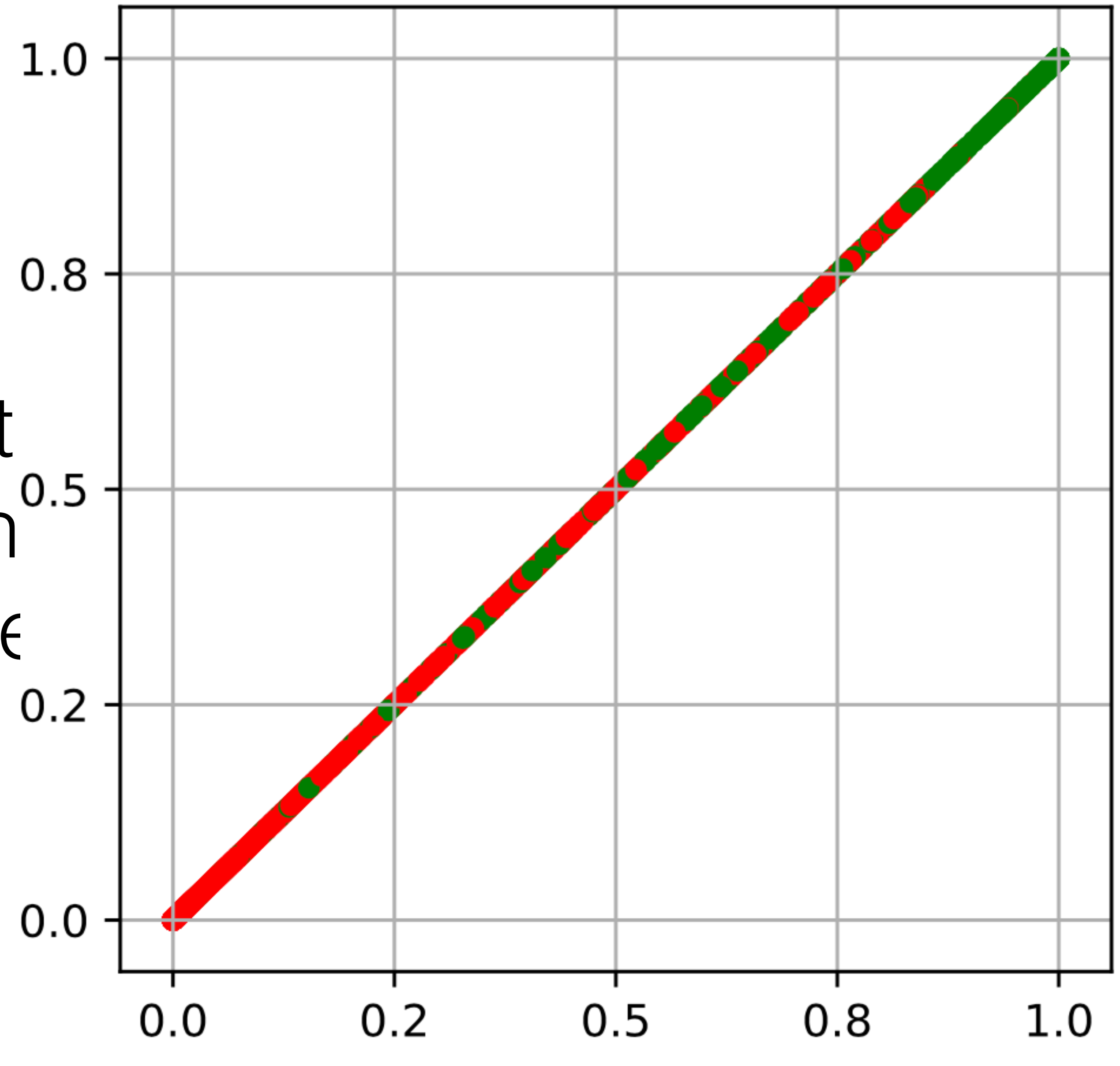


Motivation

- We want to enforce the condition that the confidence of the coarse category should be distributed among its ‘fine’ categories.

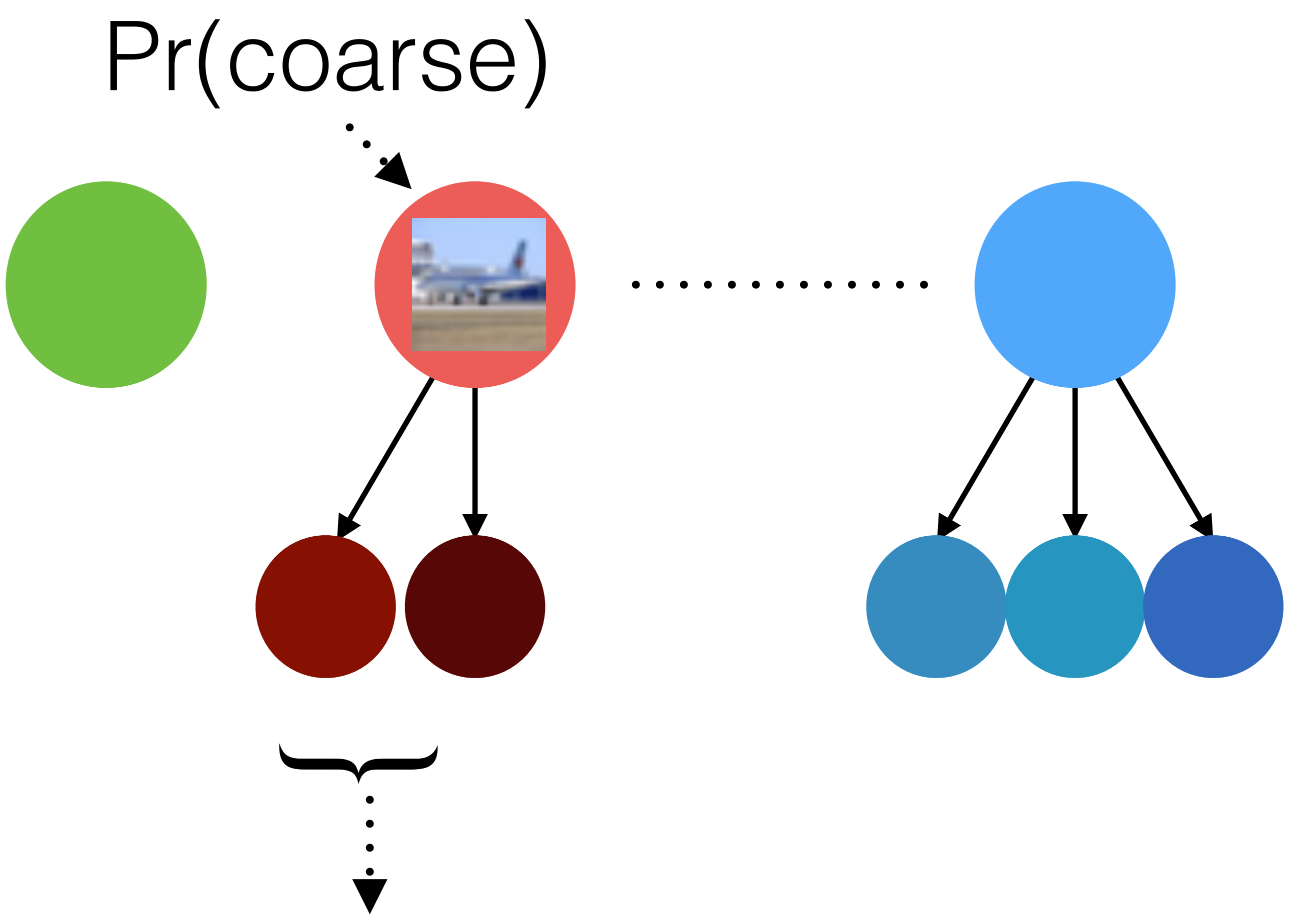
Motivation

- We want ... the confiden... ould be distribute...



Fine and Coarse Labels

Fine and Coarse Labels



$\sum \Pr(\text{fine}) = \Pr(\text{coarse})$

Method

Method

- Obtain probability values for test image from coarse and fine models.

Method

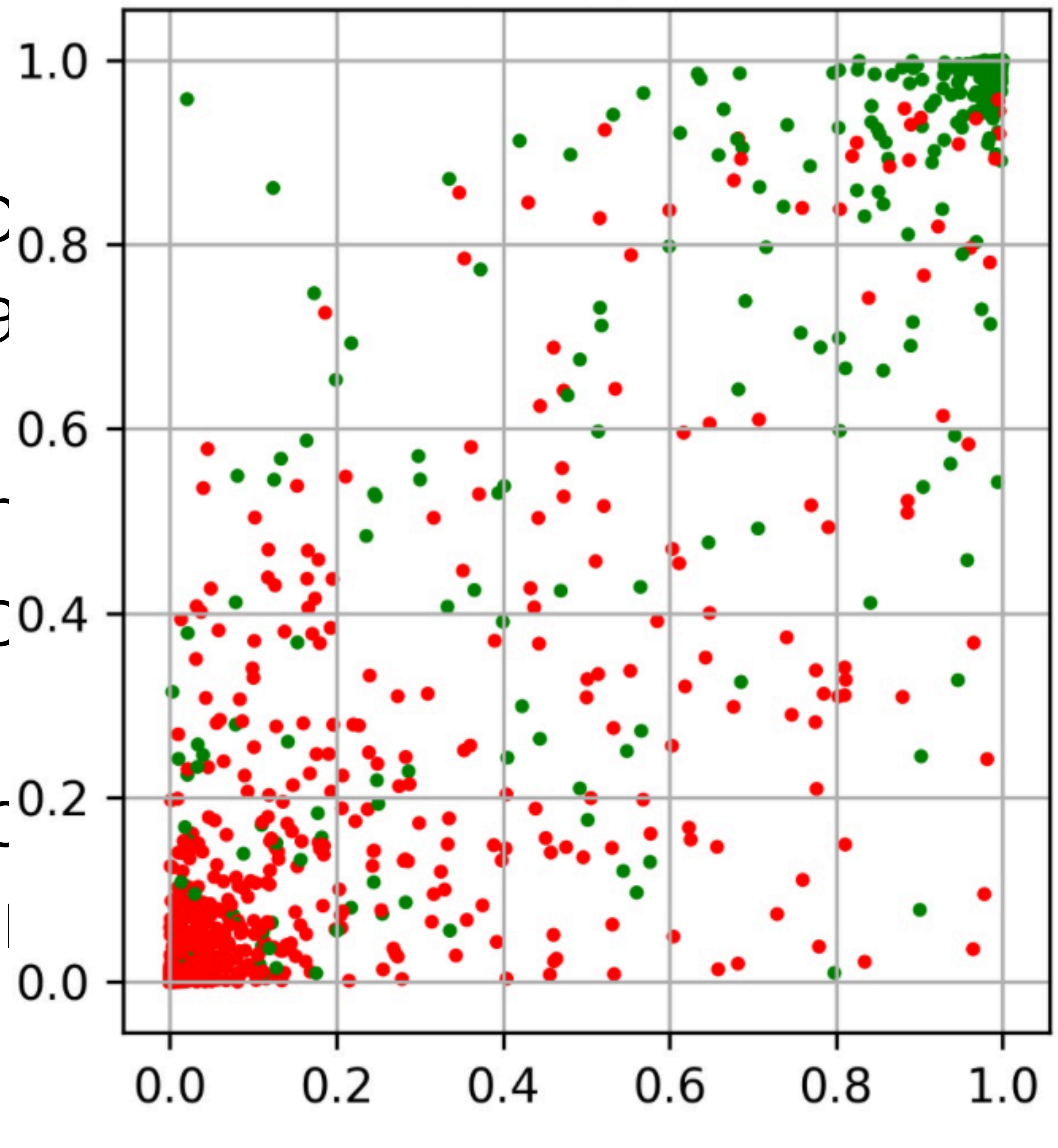
- Obtain probability values for test image from coarse and fine models.
- For each coarse category, adjust classifier prediction values such that: $\sum \Pr(\text{fine}) = \Pr(\text{coarse})$

Method

- Obtain probability values for test image from coarse and fine models.
- For each coarse category, adjust classifier prediction values such that: $\sum \Pr(\text{fine}) = \Pr(\text{coarse})$
- Predict category which has maximum probability (confidence).

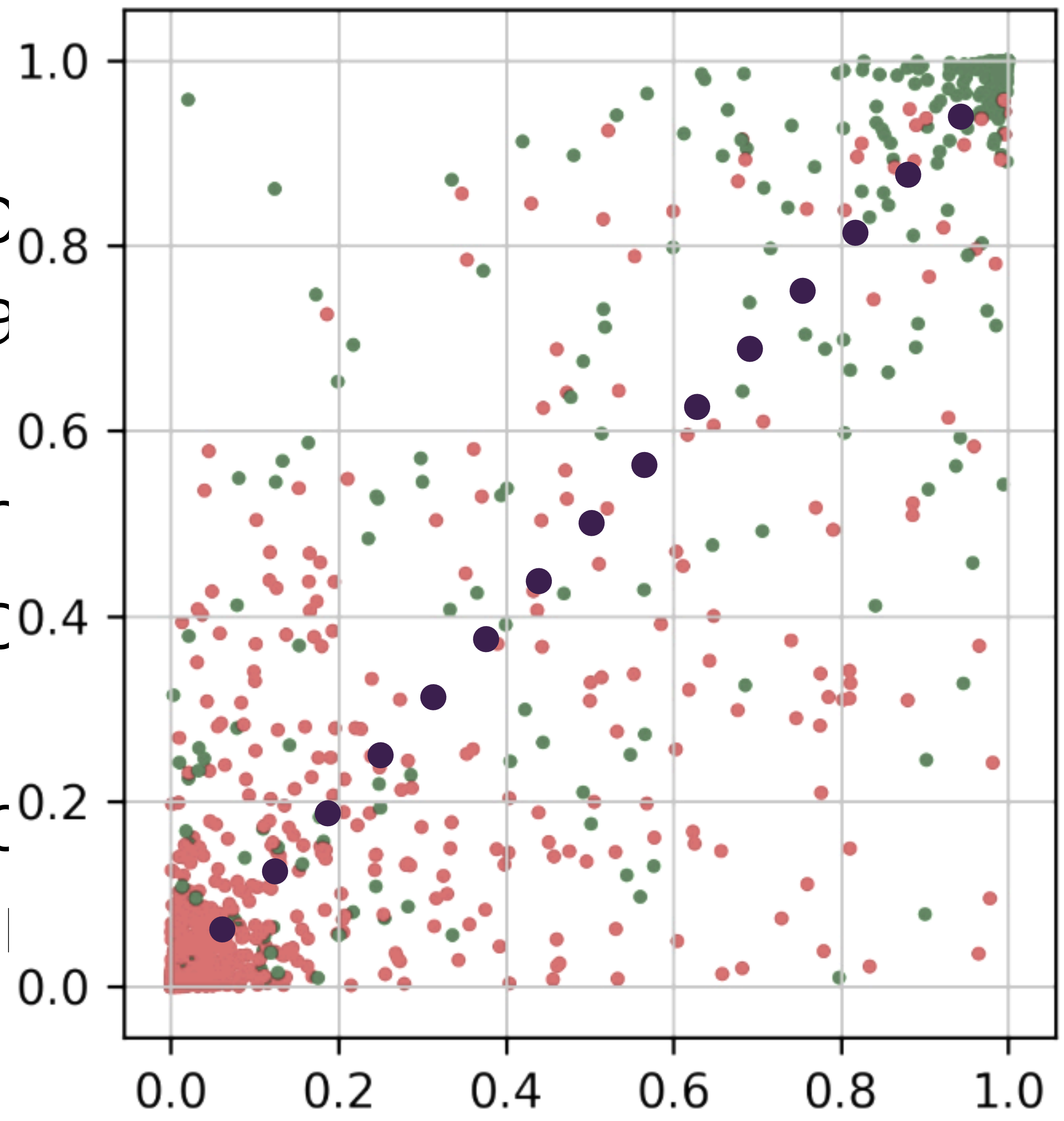
Method

- Obtain p... ge from coarse a...
- For each ... ssifier predictio...) = Pr(coarse)
- Predict c... n probability (confide...



Method

- Obtain p ge from coarse a



- For each ssifier predictic) = Pr(coarse)

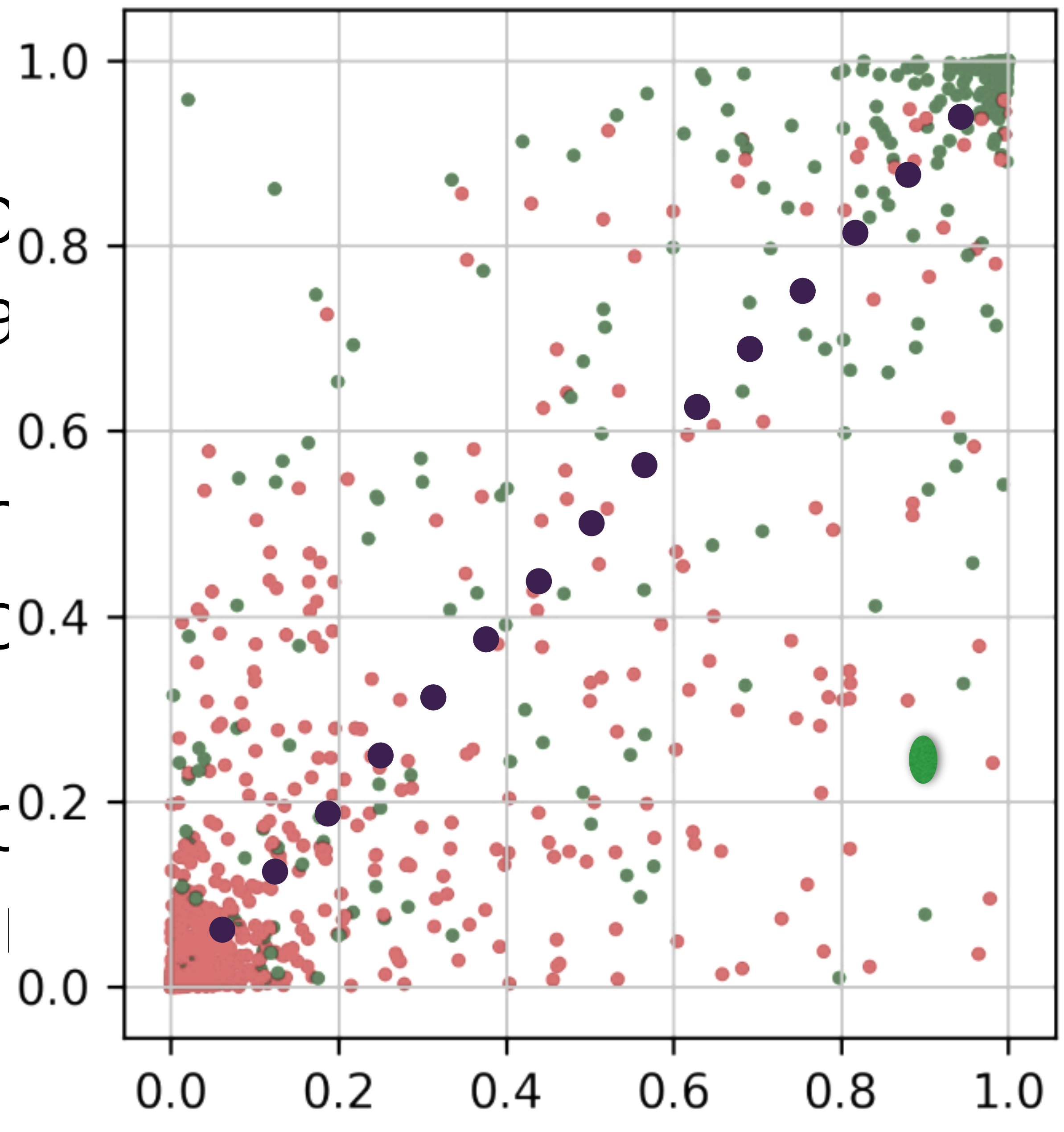
- Predict c ı probability (confide

Method

- Obtain p
coarse a

- For each
predictic

- Predict c
(confide



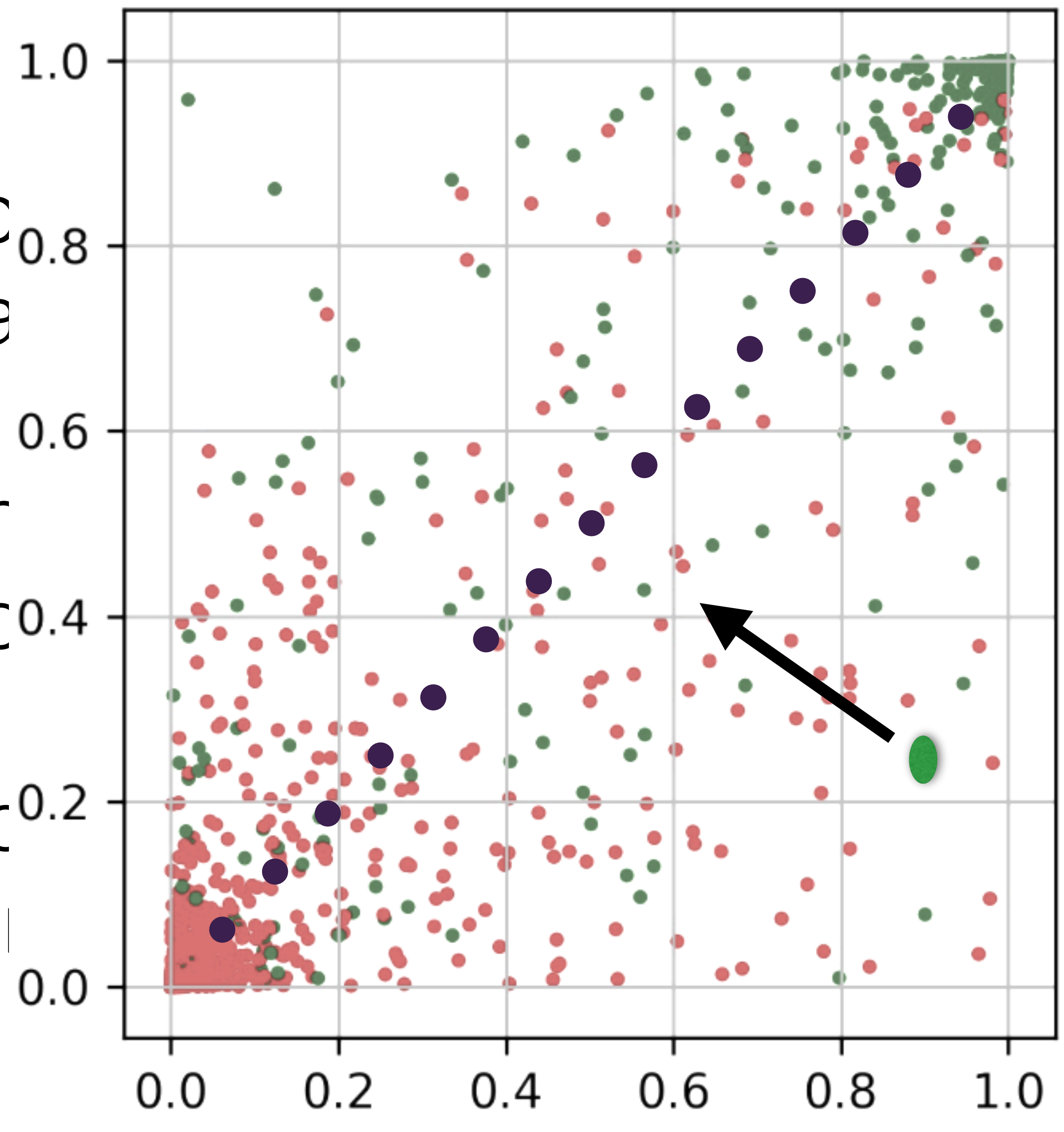
ge from

ssifier
) = Pr(coarse)

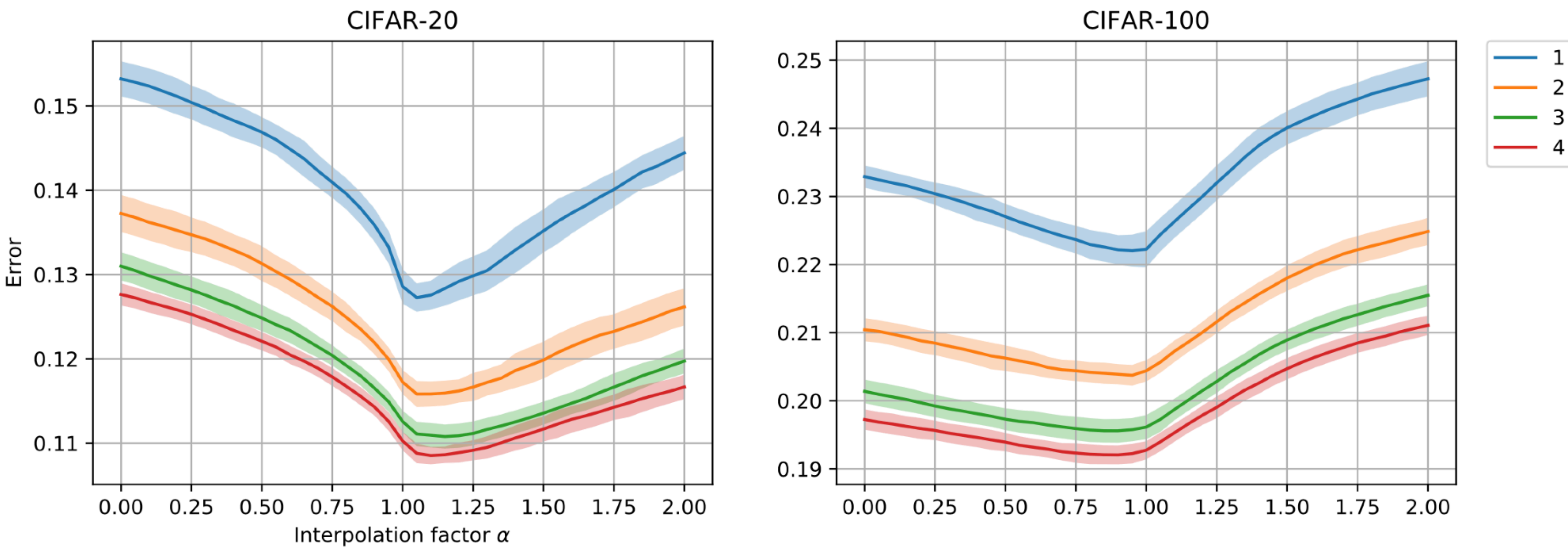
n probability

Method

- Obtain p… …ge from coarse a…
- For each… …ssifier predictic… …) = Pr(coarse)
- Predict c… …n probability (confide…



Results



- Error rate for the ResNet model after applying our adjustment scheme.
- The x-axis shows how the error changes when we move from the initial probability values to the ones enforced by our scheme.
- We observe that we steadily lower the error rate.

Method

Method

- No knowledge of ground truth categories required.

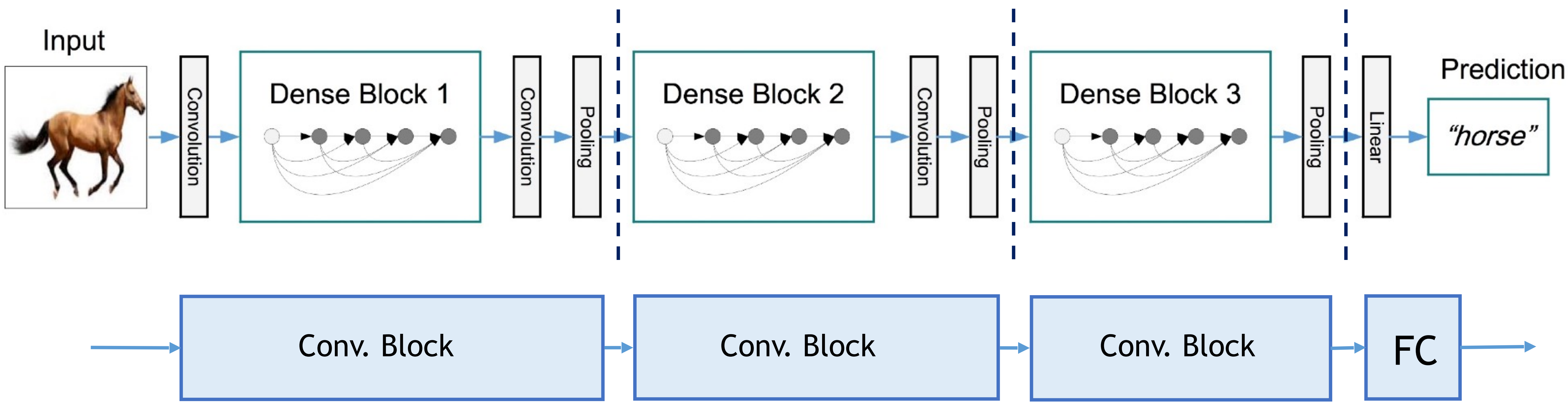
Method

- No knowledge of ground truth categories required.
- Need to know only semantic relationship among categories.

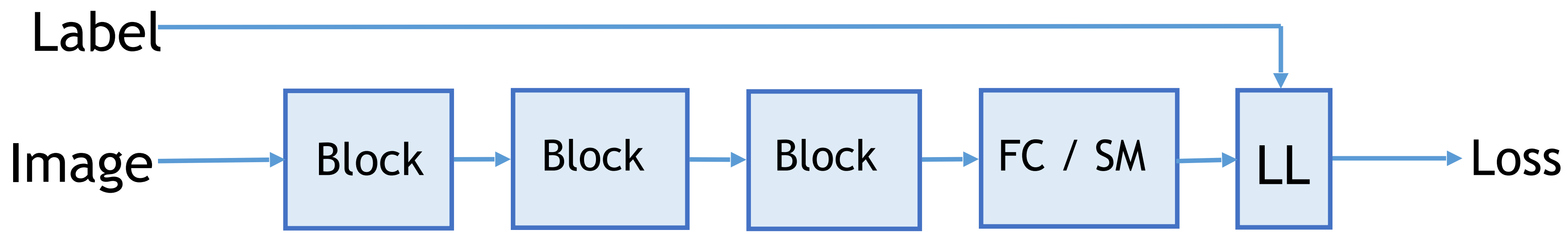
Method

- No knowledge of ground truth categories required.
- Need to know only semantic relationship among categories.
- Can be applied to any classifier which predicts probabilities over categories.

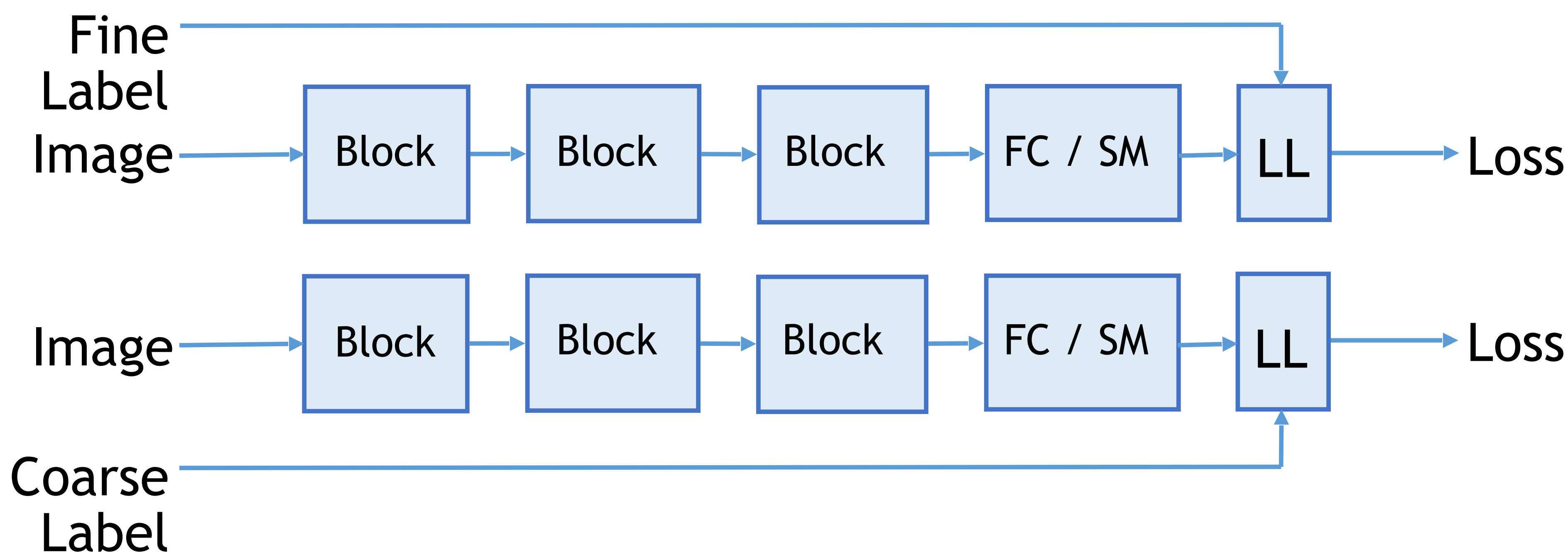
Merged Architecture



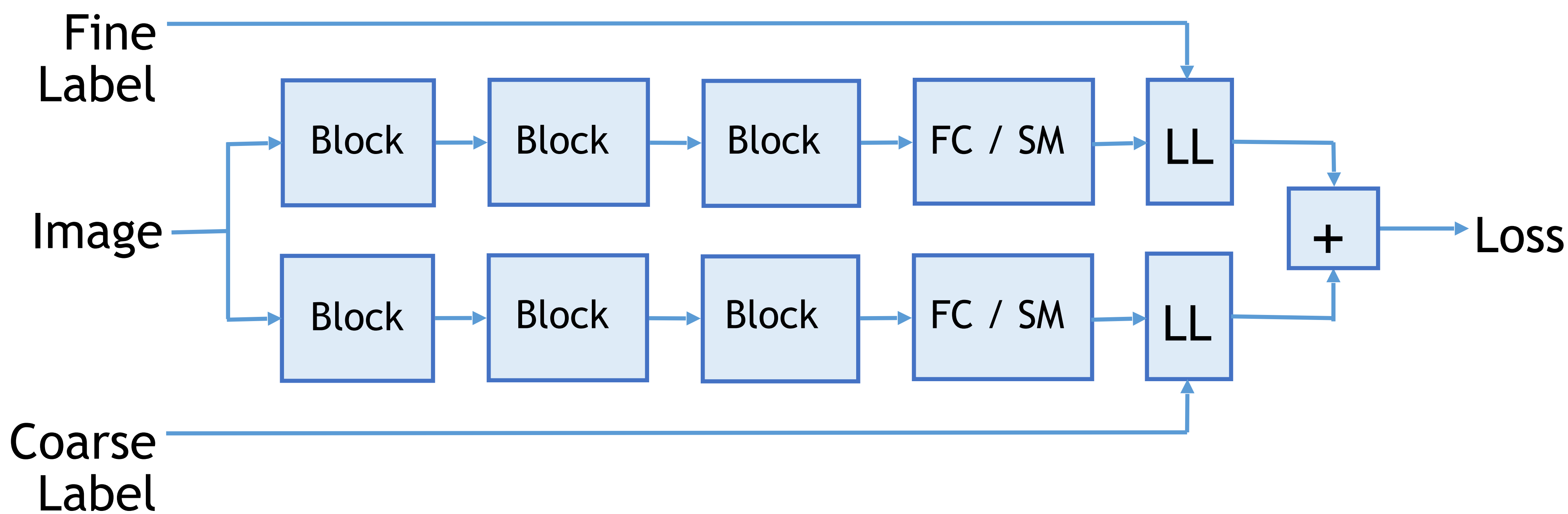
Merged Architecture



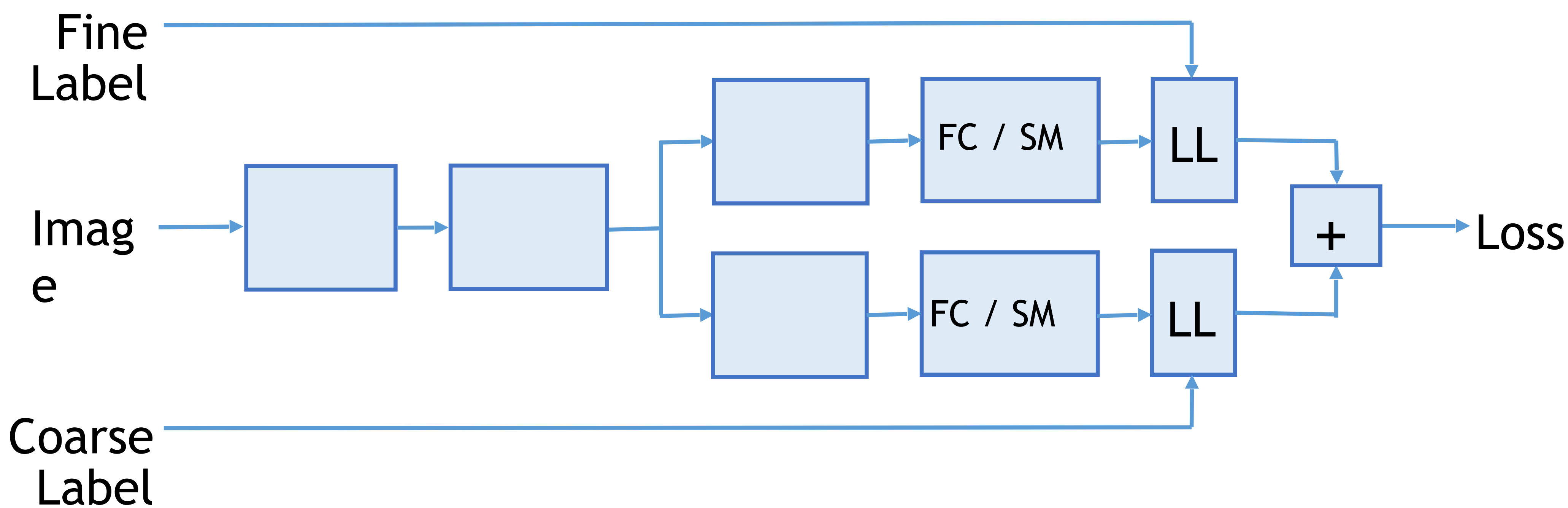
Merged Architecture

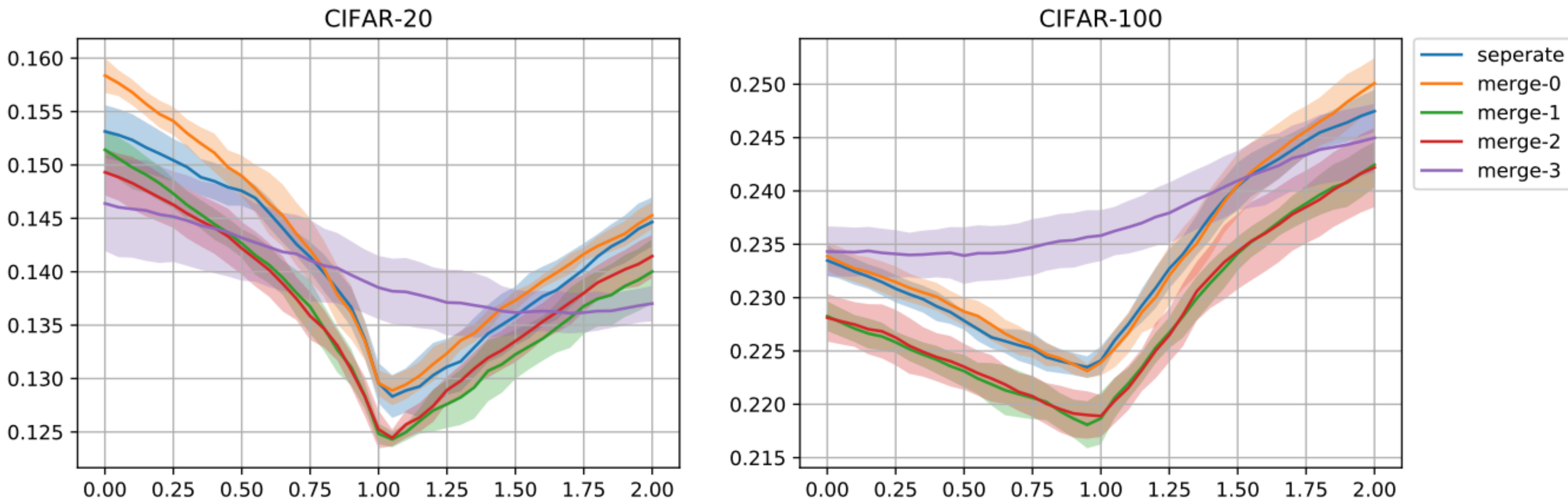


Merged Architecture



Merged Architecture




CIFAR-20
0.160
0.155
0.150
0.145
0.140
0.135
0.130
0.125
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00
CIFAR-100
0.250
0.245
0.240
0.235
0.230
0.225
0.220
0.215
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75 2.00
seperate
merge-0
merge-1
merge-2
merge-3

Conclusion

Conclusion

- Simple method which does not require any training.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.
- Significant performance improvement over both the fine and coarse grain classification by simply using the generic-specific information.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.
- Significant performance improvement over both the fine and coarse grain classification by simply using the generic-specific information.
- Future work:

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.
- Significant performance improvement over both the fine and coarse grain classification by simply using the generic-specific information.
- Future work:
  - Penalize difference between fine and coarse category probabilities during training phase.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.
- Significant performance improvement over both the fine and coarse grain classification by simply using the generic-specific information.
- Future work:
  - Penalize difference between fine and coarse category probabilities during training phase.
  - Apply to muti-level hierarchies.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.
- Significant performance improvement over both the fine and coarse grain classification by simply using the generic-specific information.
- Future work:
  - Penalize difference between fine and coarse category probabilities during training phase.
  - Apply to muti-level hierarchies.
  - Apply other relations among probabilities at different levels.

Conclusion

- Simple method which does not require any training.
- Can be applied to ***any*** trained model.
- Only in the case of mutually exclusive and complete classes and in the case of two levels of hierarchy.
- Significant performance improvement over both the fine and coarse grain classification by simply using the generic-specific information.
- Future work:
  - Penalize difference between fine and coarse category probabilities during training phase.
  - Apply to muti-level hierarchies.
  - Apply other relations among probabilities at different levels.
- To appear in: *ACM International Conference on Multimedia Retrieval (ICMR), June 2017.*

Thank You.