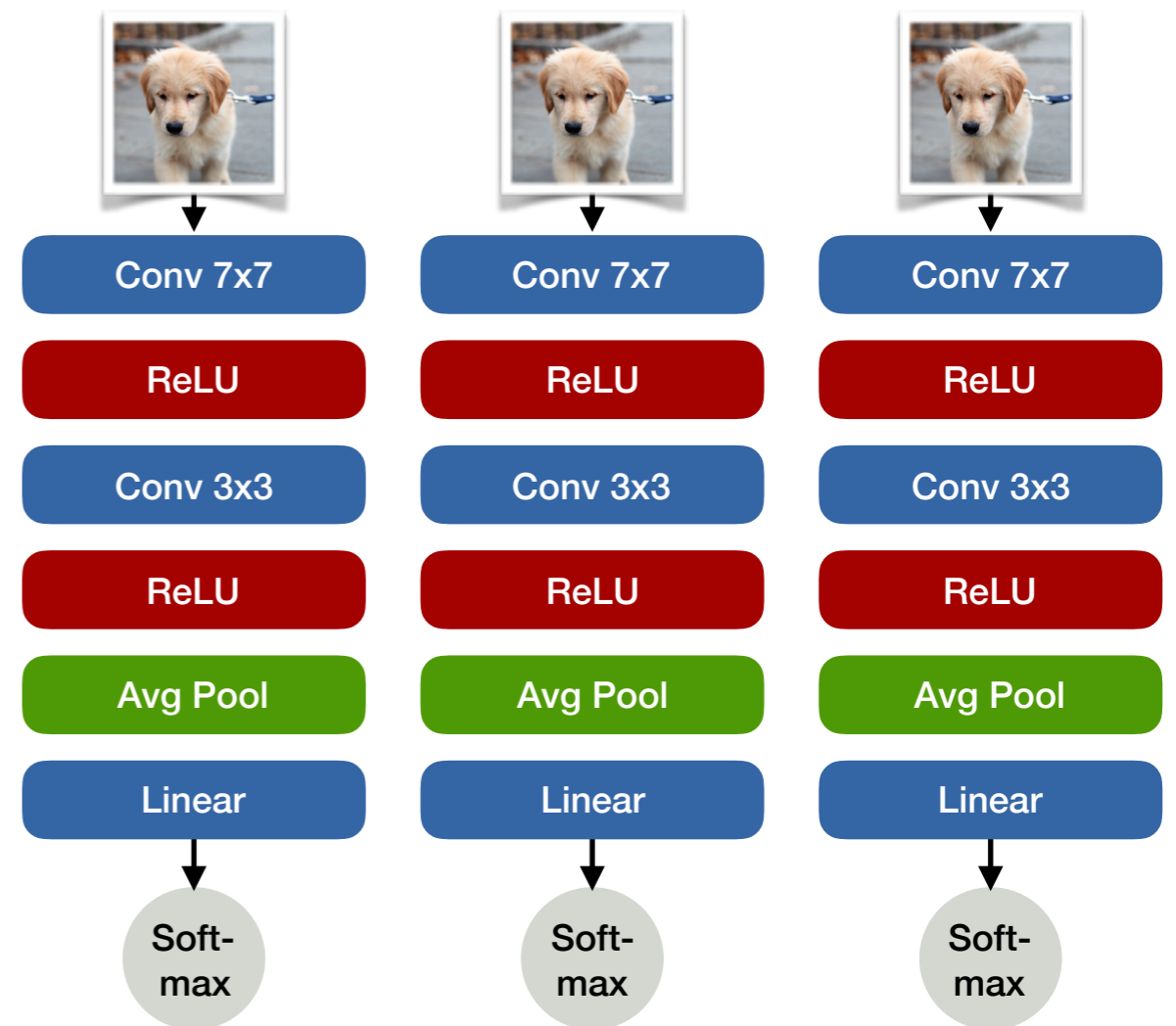


# Ensembles

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

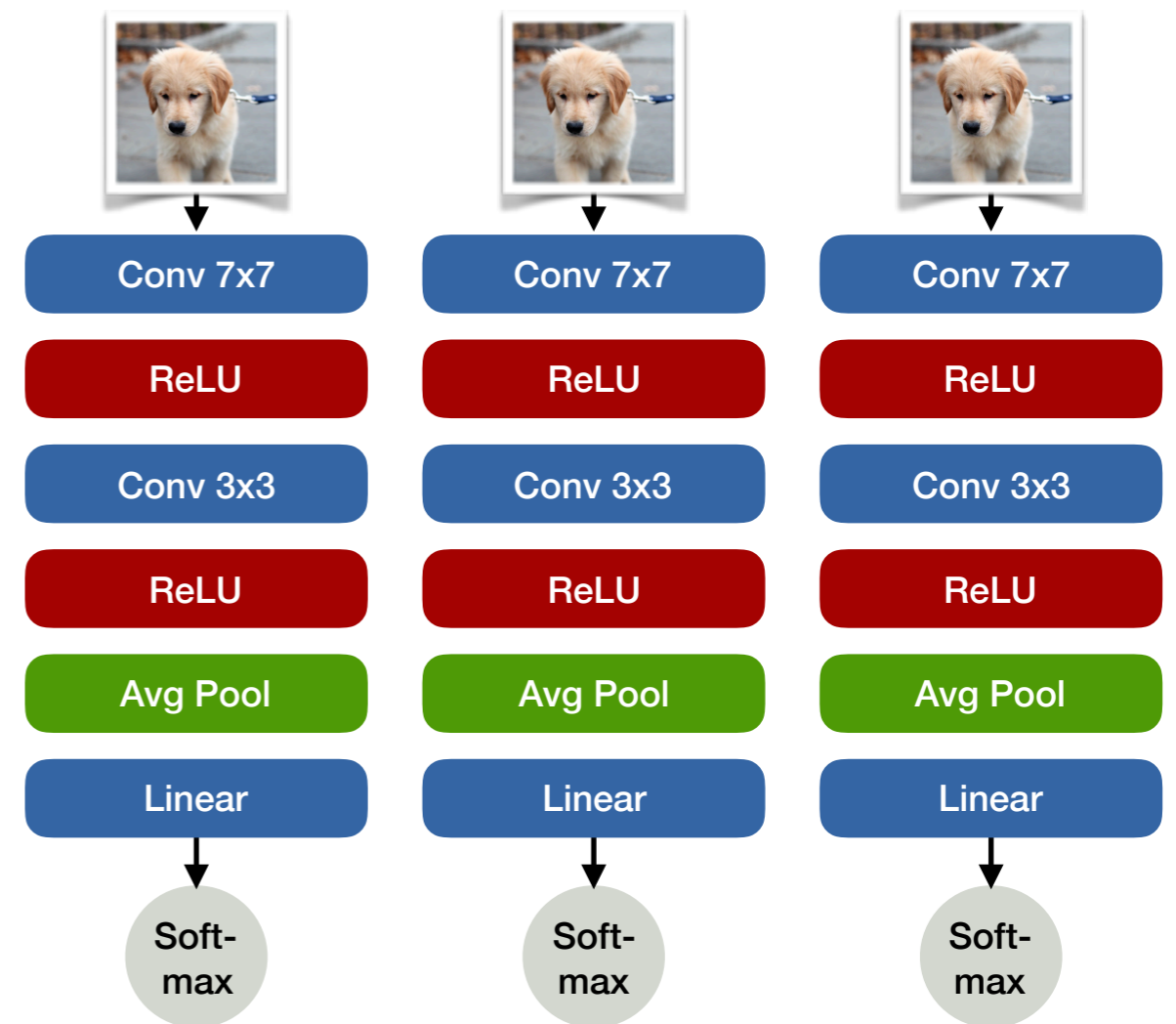
# Ensembles

- Train multiple models
- Average predictions of multiple models



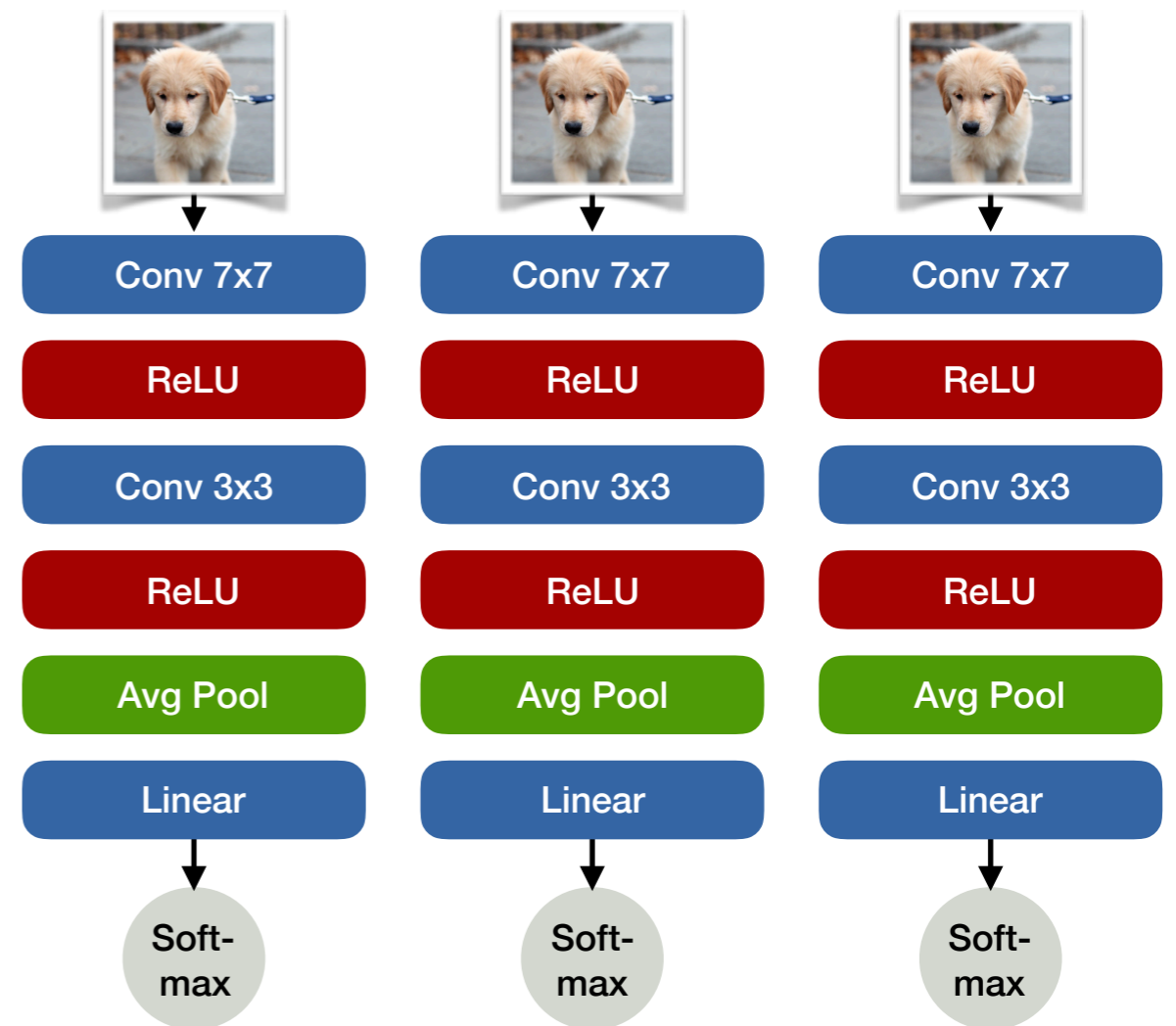
# Ensembles

- Pre-deep learning
  - Use different subsets of training data
- Deep learning
  - Use different random initializations / data augmentation
  - Different local minima



# Why do ensembles work?

- Fewer parameters / model
- Each model overfits in its own way
- Usually a 1-3% accuracy boost on most tasks
- longer training



# Why do we average predictions?

- For a convex loss function
- loss of average prediction  $<$  average loss of individual models

# When to use ensembles?

- If you have the compute power
- If you really need the last bit of accuracy
  - e.g. production, competitions