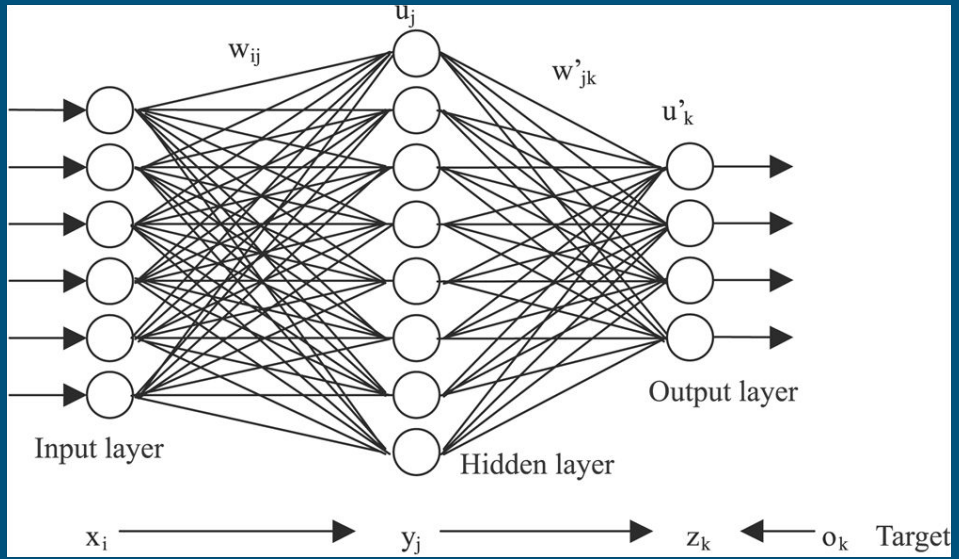
Automatic Classification of Large Biomedical Data

Tom Corcoran

Approach

- Artificial Neural Networks (ANN)
- Feature Extraction
- Training
- Classification
- Colorization



Current State

- Existing Work by Larsson, Maire, Shakhnarovic
- Small-scale
- One NVIDIA Titan X GPU
- 17.5 Hour Training Time
- Python
- Significant number of Failure Modes

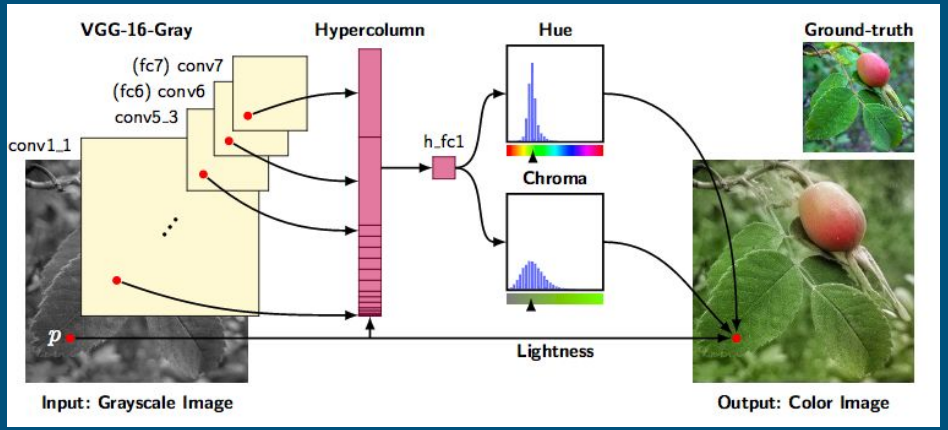


Image source: http://people.cs.uchicago.edu/~larsson/colorization/

Successes

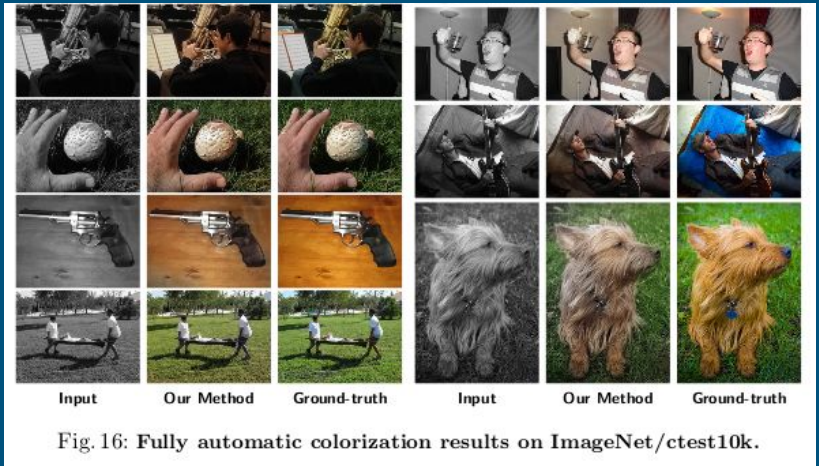


Fig. 16: **Fully automatic colorization results on ImageNet/ctest10k.**

Failure Modes

Source for images: http://arxiv.org/pdf/1603.06668.pdf

My Tasks

- Develop an extensible 3-tier framework for development
- ANN => MPI+CUDA => App-Layer / Filter
- Rewrite existing Python code base in C to allow for scalability and introduction of large -scale parallelism
- Utilize MPI to allow distributed memory use
- Tune the CUDA aspect / replace with OpenACC if possible

Blue Waters / HPC Component

- Scaling past 1 node
- Decreased training time
- Increased network complexity (number of hidden layers) - may or may not be a good idea
- Improve accuracy of classification
- Provide a highly-tuned ANN useful for real-world classification use