CS6501: Deep Learning for Visual Recognition

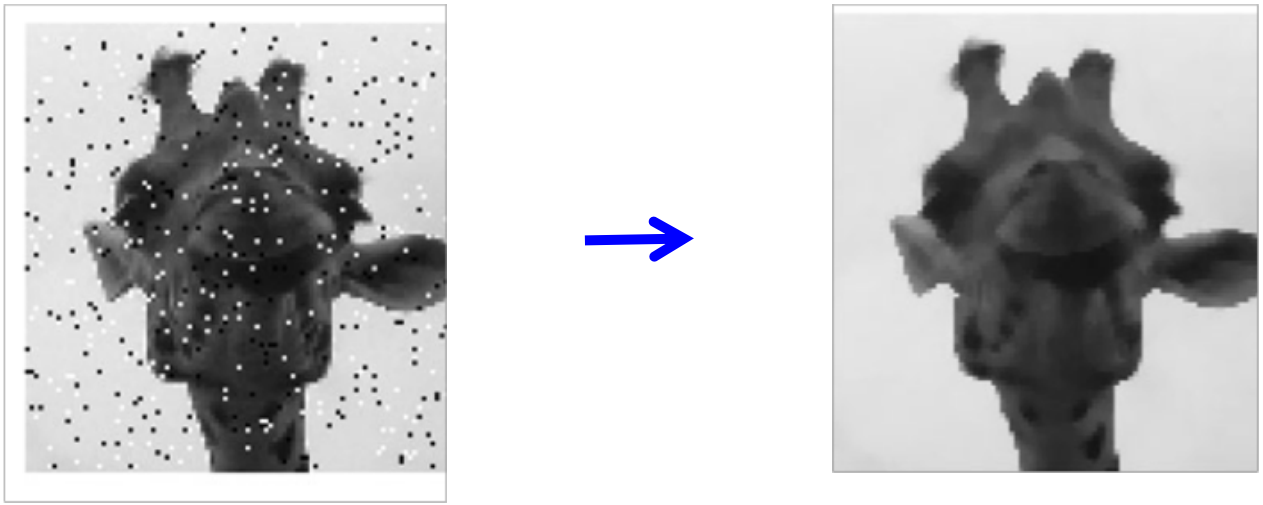
Course Recap and What's Next?

Course Recap

- What did we learn this semester?

Image Processing and Manipulation

- Image Processing – Brightness, Saturation, Color Spaces
- Image Convolution / Filtering: Blurring, Sharpness, Noise Removal
- Image Gradients / Sobel Operators / Features



Basic Machine Learning and Image Classification

- Softmax Classifier / Linear Classifiers in General
- Stochastic Gradient Descent (SGD)
- Momentum, Regularization, Choosing Hyperparameters

Deep Learning

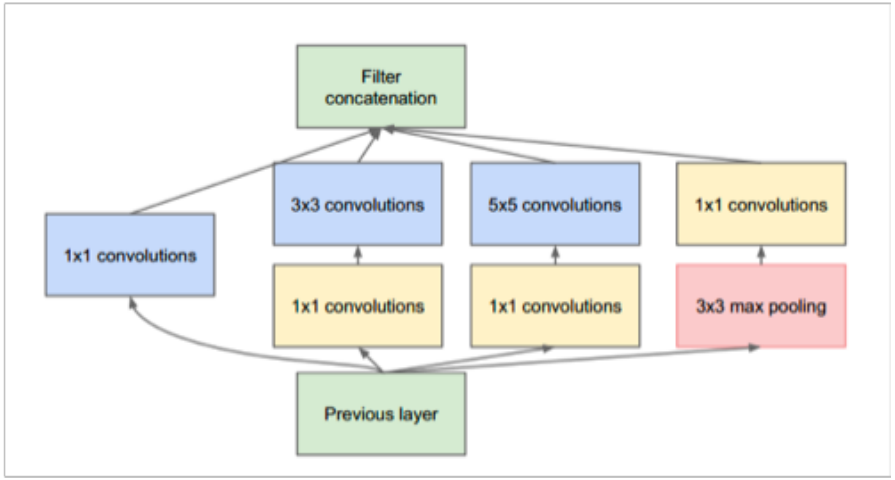
- Multi-layer Perceptron / Neural Networks
- The backpropagation algorithm
- Implementing our own Deep Learning Toolkit / Library

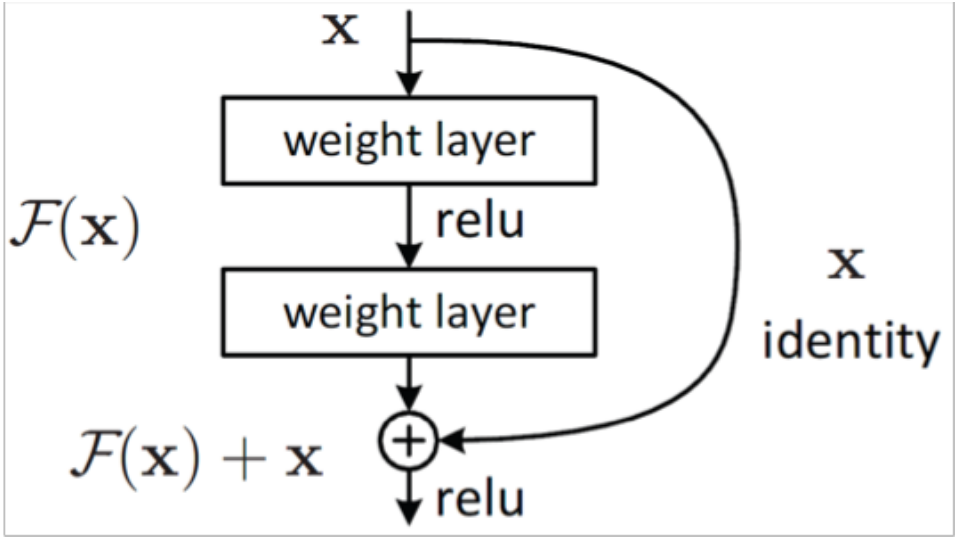
$$a_1 = sigmoid(w_{[1]}x^T + b_{[1]}^T)$$
$$a_2 = sigmoid(w_{[2]}a_1^T + b_{[2]}^T)$$
$$\dots$$
$$a_i = sigmoid(w_{[k]}a_{k-1}^T + b_{[k]}^T)$$
$$\dots$$
$$f = softmax(w_{[n]}a_{n-1}^T + b_{[n]}^T)$$
$$l = loss(f, y)$$

$$\frac{\partial l}{\partial w_{[k]ij}} = \frac{\partial l}{\partial a_{n-1}} \frac{\partial a_{n-1}}{\partial a_{n-2}} \dots \frac{\partial a_k}{\partial a_{k-1}} \frac{\partial a_{k-1}}{\partial w_{[k]ij}}$$

Deep Learning for Visual Recognition: Classification

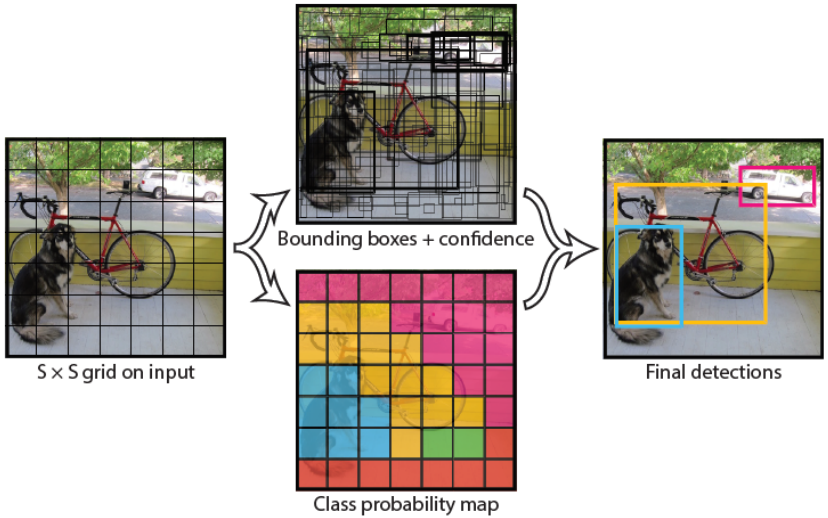
- Alexnet, VGG Networks.
- Residual Connections, Resnets
- Google Net, Inception Nets

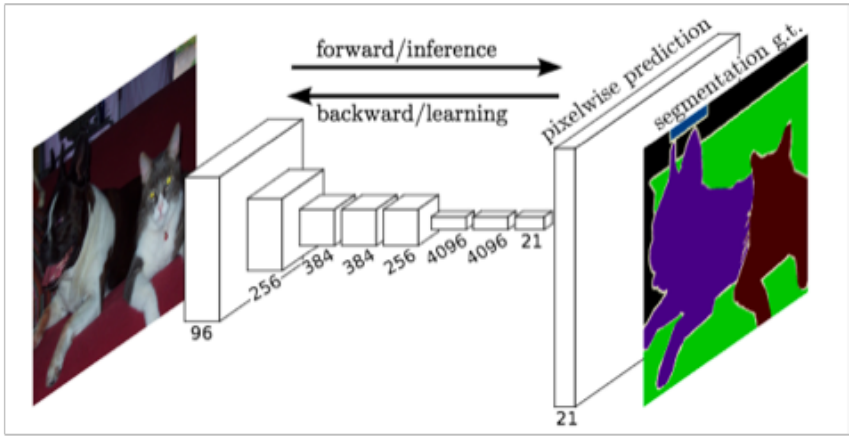




Deep Learning for Visual Recognition: Detection / Segmentation

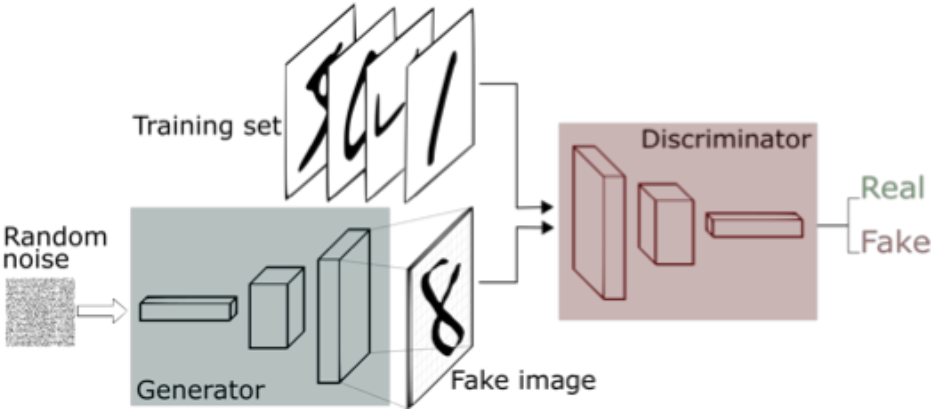
- Convolutional Neural Networks (CNNs)
- Object Detection: RCNN, Fast-RCNN, Faster-RCNN, SSD, YOLO, SNIPER
- Image Segmentation: Deconvolutions, Fully Convolutional Networks, U-Nets.

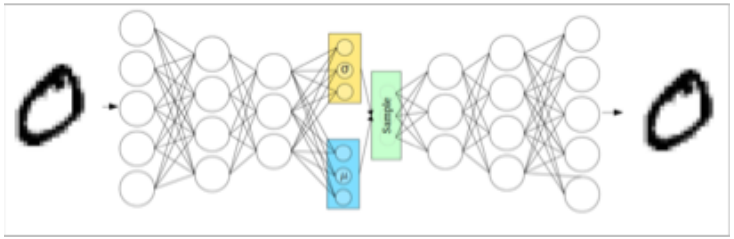




Deep Learning for Visual Recognition: Synthesis

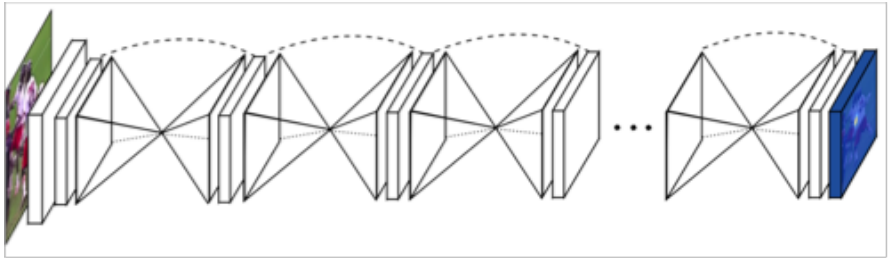
- Generating Images: Generative Adversarial Networks (GANs), Variational Auto-Encoder (VAEs)

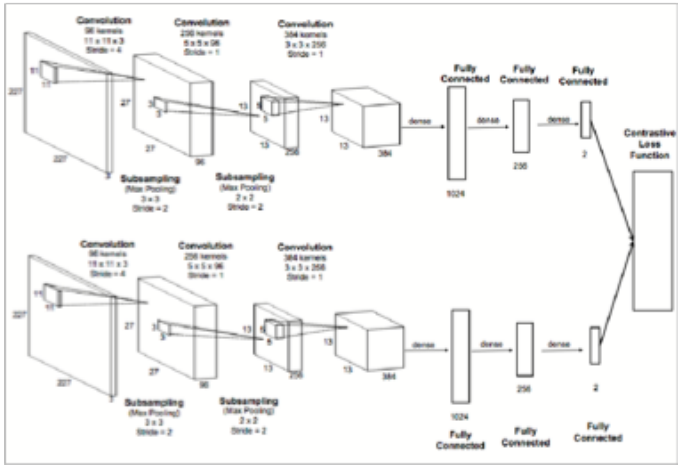




Deep Learning for Visual Recognition: Other Tasks

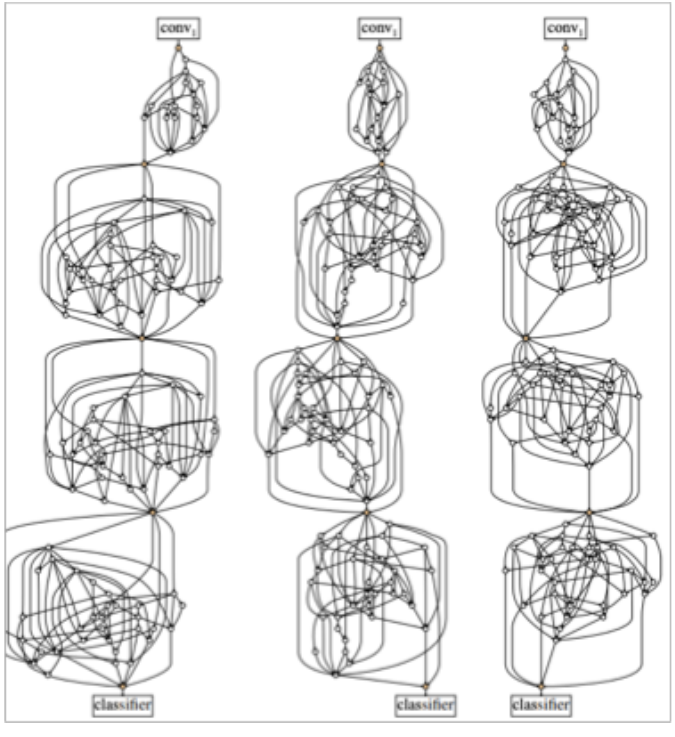
- Human Pose Estimation: Hourglass Networks
- Matching: Siamese Networks, Triplet Losses





What else is coming? Image Classification

- Models have almost peaked on Imagenet
- Model search: How to find model architectures automatically!

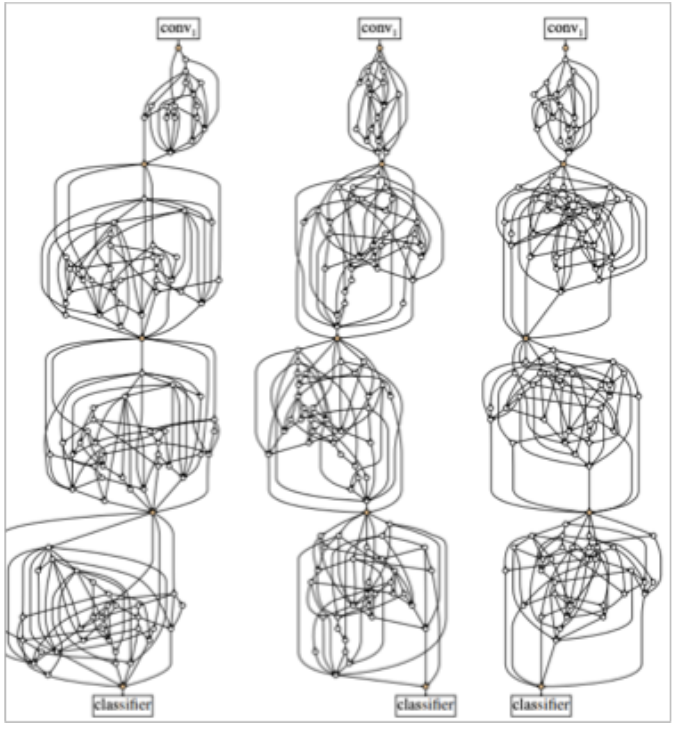


Saining Xie Alexander Kirillov Ross Girshick Kaiming He

Exploring Randomly Wired Neural Networks for Image Recognition

What else is coming? Image Classification

- Models have almost peaked on Imagenet
- Model search: How to find model architectures automatically!



These models are as fast as Resnet models but ~2% more accurate (79% vs 77%)

Saining Xie Alexander Kirillov Ross Girshick Kaiming He

Exploring Randomly Wired Neural Networks for Image Recognition

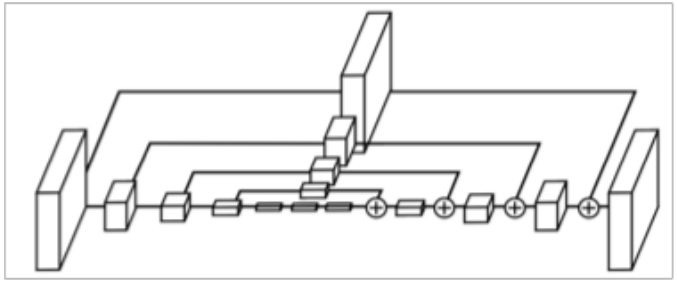
What else is coming? Object Detection

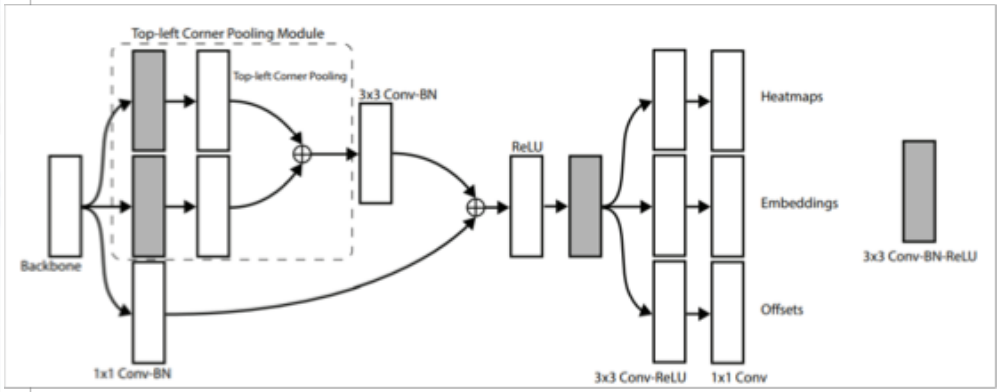
- CornerNets, CenterNets

"CornerNet: Detecting Objects as Paired Keypoints" by Hei Law, and Jia Deng ECCV 2018

What else is coming? Object Detection

- CornerNets, CenterNets

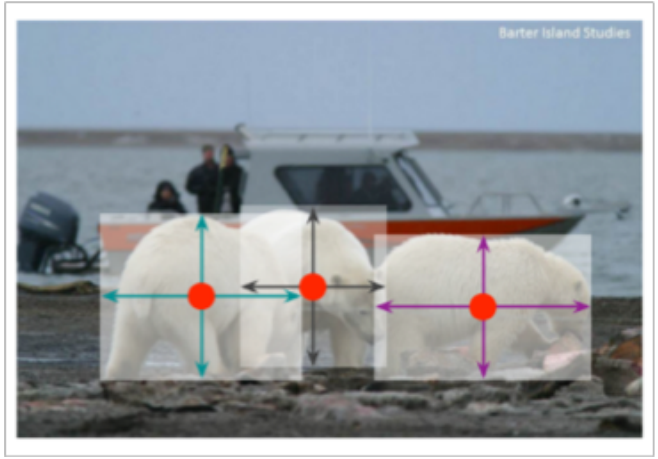




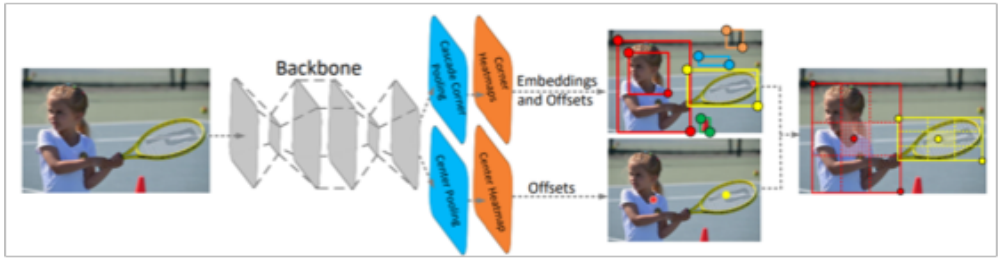
"CornerNet: Detecting Objects as Paired Keypoints" by Hei Law, and Jia Deng ECCV 2018

What else is coming? Object Detection

- CornerNets, CenterNets

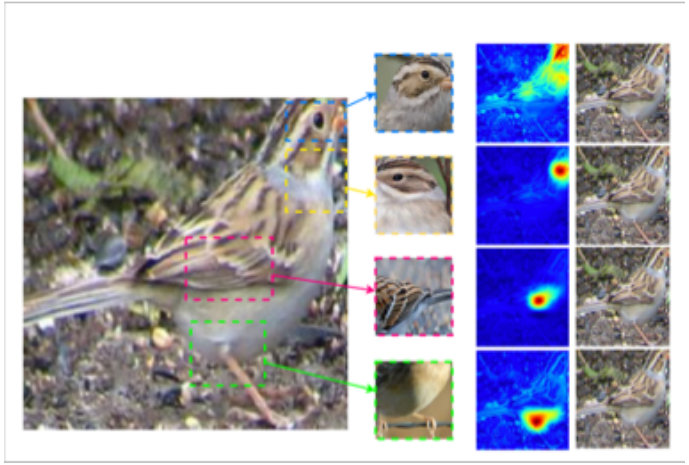


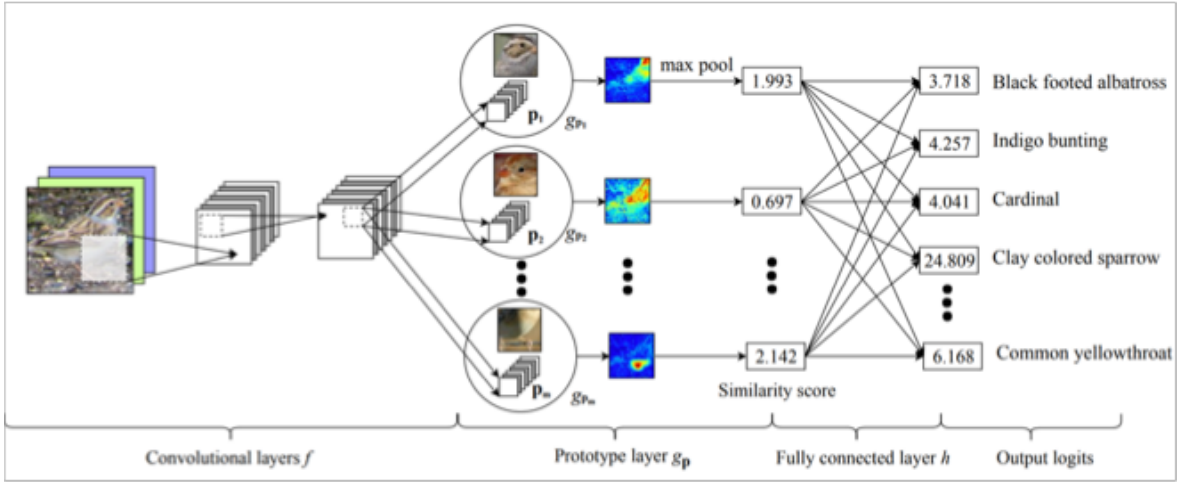
Objects as Points. Xingyi Zhou, Dequan Wang, Philipp Krähenbühl. April 2019.



*CenterNet: Keypoint Triplets for Object Detection by Kaiwen Duan et al. April 17, 2019 arXiv

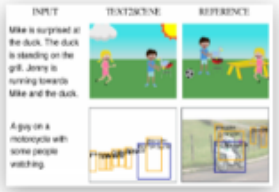
Interpretable Deep Learning



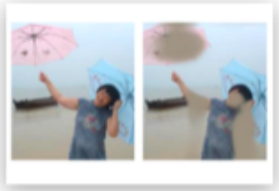


This Looks Like That: Deep Learning for Interpretable Image Recognition. Chen et al 2018.

Here at UVA



NEW! Text2Scene: Generating Compositional Scenes from Textual Descriptions
Fuwen Tan, Song Feng, Vicente Ordonez.
Intl. Conference on Computer Vision and Pattern Recognition. **CVPR 2019**.
Long Beach, California. June 2019. [arxiv] [bibtex] *(~Oral presentation)*



NEW! Adversarial Removal of Gender from Deep Image Representations
Tianlu Wang, Jieyu Zhao, Mark Yatskar, Kai-Wei Chang, Vicente Ordonez.
arXiv:1811.08489. November 2018.
[arxiv] [bibtex] In Submission

Here at UVA

- How to learn Interpretable and Robust models?
- How to learn models that do not need as much training data?
- How to learn faster models? e.g. XNOR-Nets
- How to better represent multimodal features? (images + text)
- How to build compositional models for unusual images?

Thanks and All the Best!