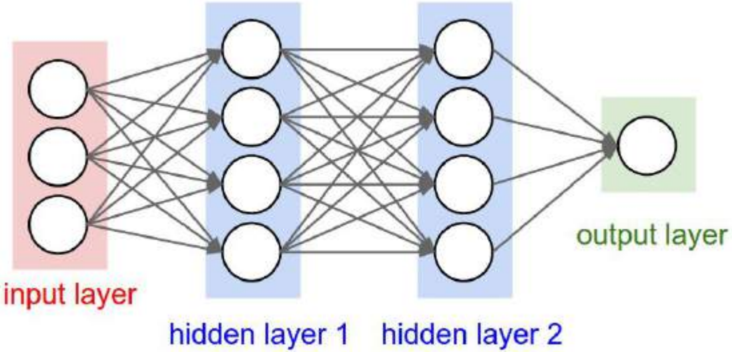
NEURAL RENDERING MODEL (NRM): JOINT GENERATION AND PREDICTION FOR SEMI-SUPERVISED LEARNING

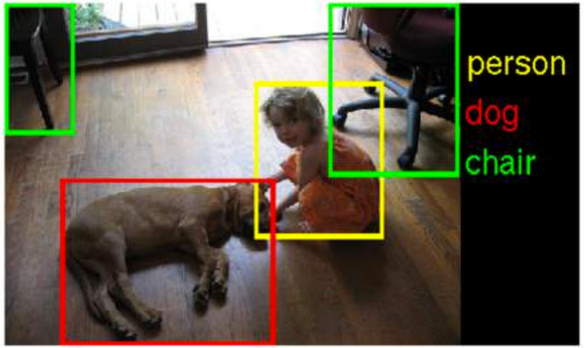
Tan Nguyen

Joint work with Nhat Ho, Ankit B. Patel,
Anima Anandkumar, Michael I. Jordan, and Richard G. Baraniuk



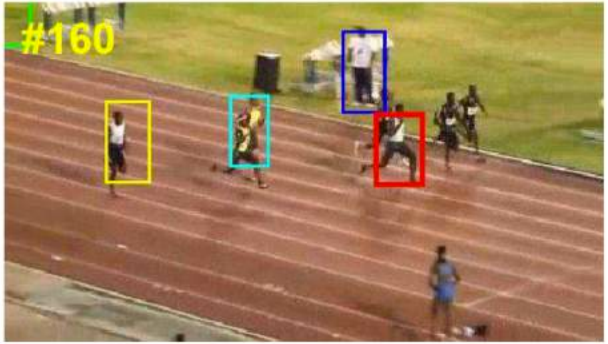
RICE

Deep Learning – The Current State-of-the-Art



Object Recognition

Image Segmentation



Object Tracking

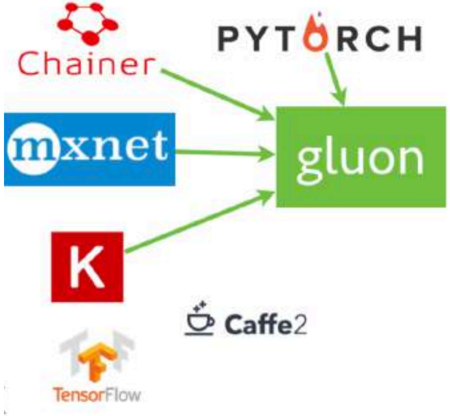
Deep learning models, such as the Deep Convolutional Networks, achieve state-of-the-art performance in various visual perception tasks

Deep Learning – Behind the Scene

Massive Amount of Labeled Data

Big Computer

Efficient Frameworks



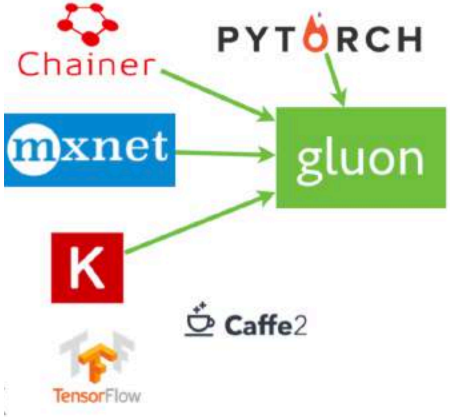
Deep Learning – Behind the Scene

Massive Amount of
Labeled Data

Big Computer

AWS,
Google Cloud,
Salesforce,
…

Efficient Frameworks



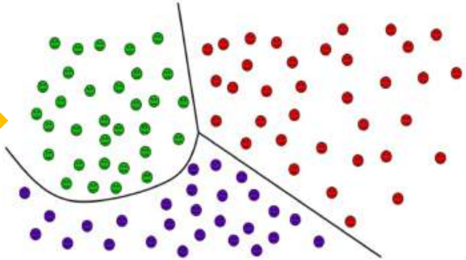
MXNet,
PyTorch,
Tensorflow,
…

Deep Learning – Behind the Scene

It is often costly or infeasible to acquire and annotate massive datasets

Semi-Supervised Learning

Raw data from HTML

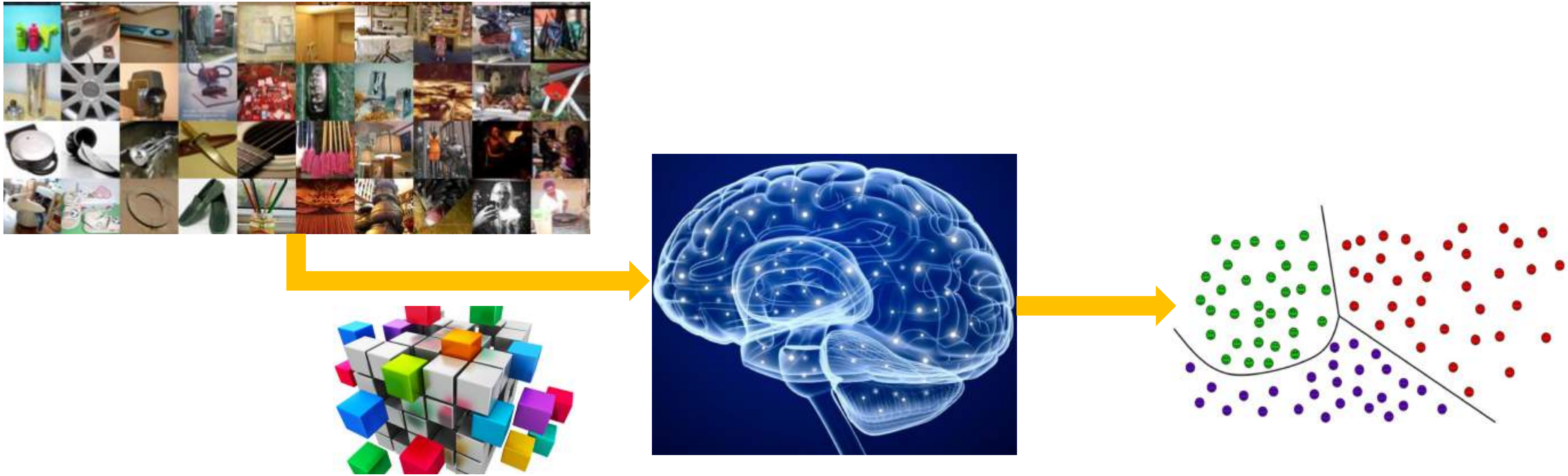


Cat

Dog

Learning from Unlabeled Data is Challenging

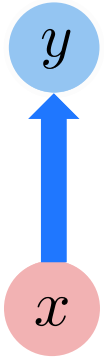
Raw data from HTML



One solution for unsupervised learning is to employ the hidden structure in the data to facilitate the learning

Predictive vs Generative Models

$P(y|x)$



Good at supervised learning

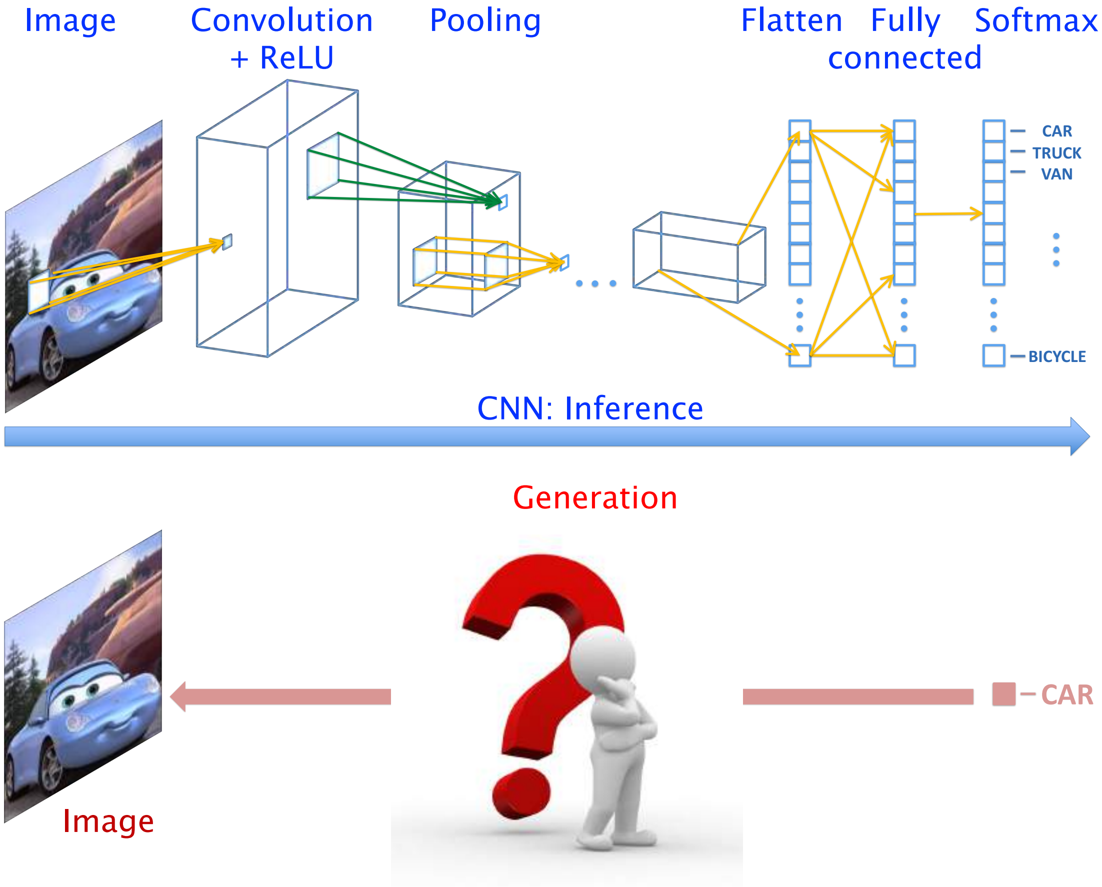
One model to do both?

$P(x|y)$

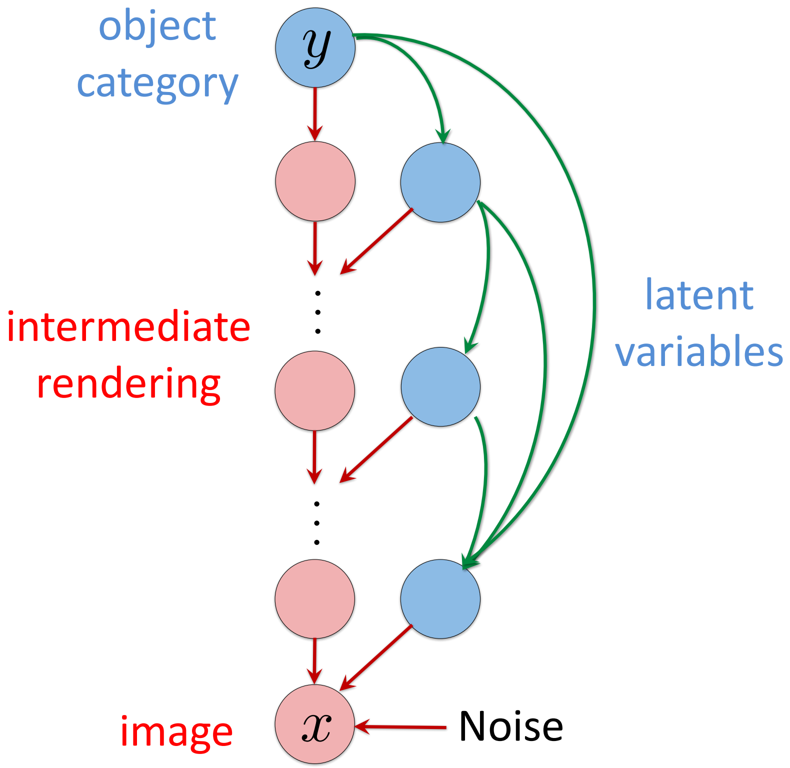
Potential solution for unsupervised learning

- CNNs are state-of-the-art predictive models
- What class of $P(x|y)$ yields CNNs for $P(y|x)$?

How can We Reverse CNNs into a Generative Model?



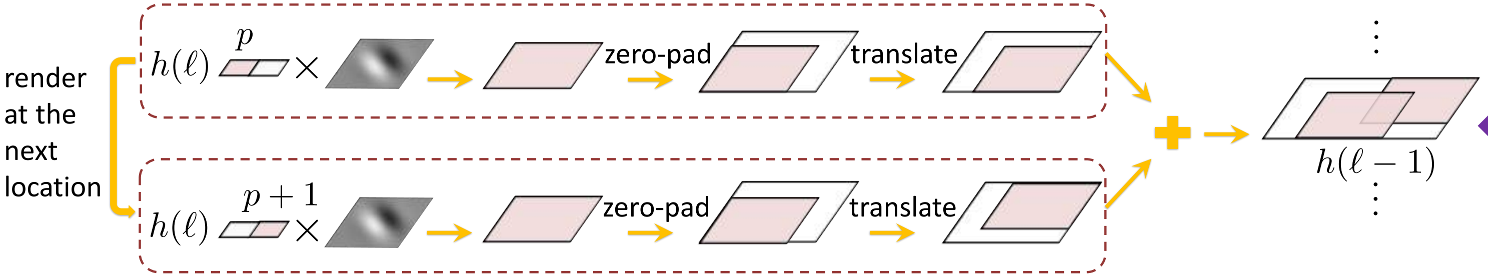
Neural Rendering Model (NRM)

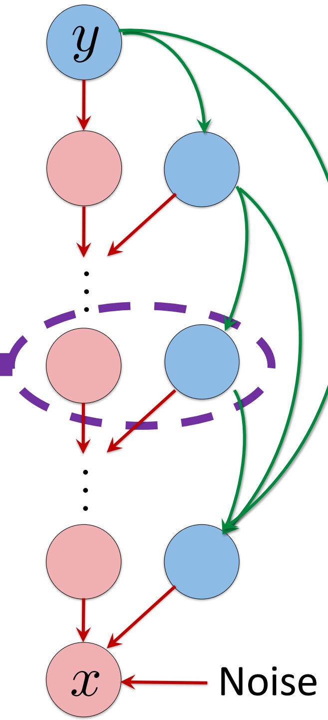


Reverse-engineer CNN to:

- Design the corresponding generative model
- Formulate the joint priors for latent variables

Rendering in One Layer of NRM

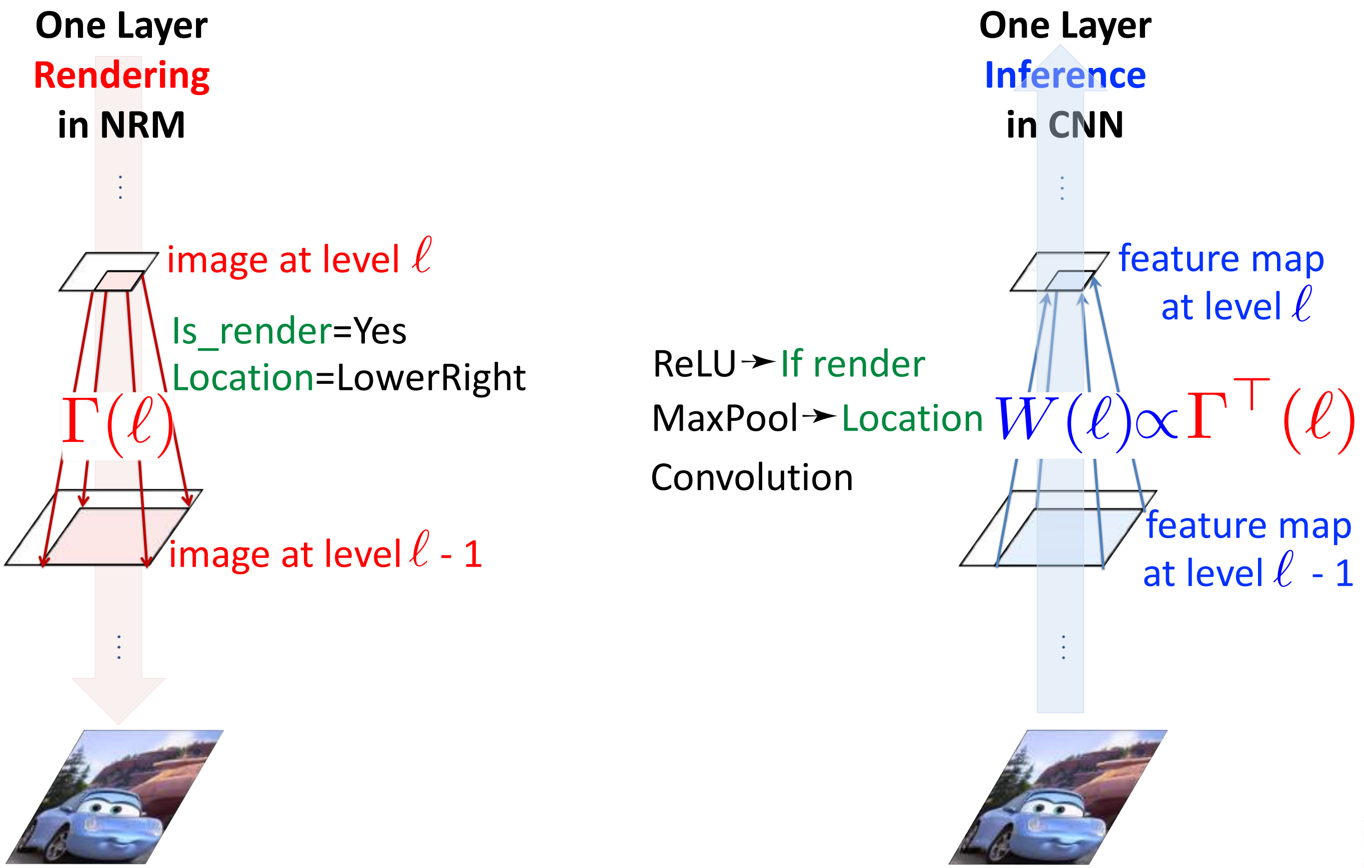




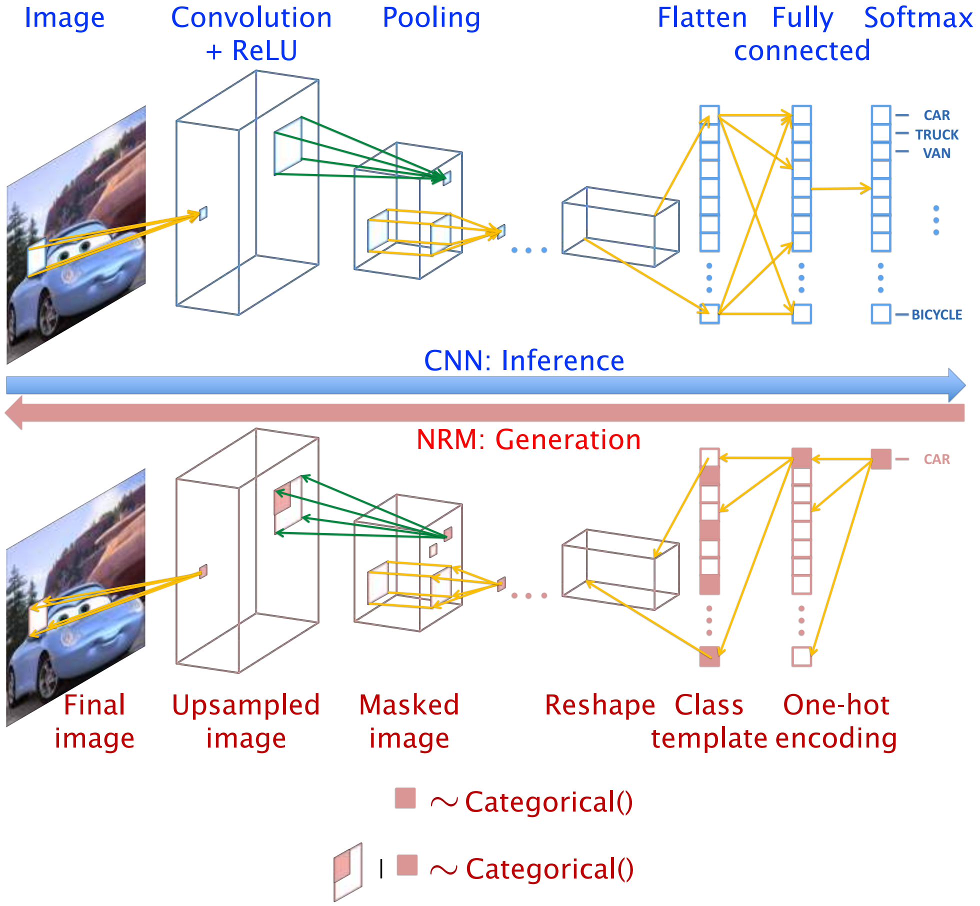
At each pixel location:

1) Decide if using that pixel to render at next layer or not
2) Zero-pad the rendered patch
3) Translate the rendered feature locally within each patch
4) Combine all patches together to form the intermediate rendered image at the next layer

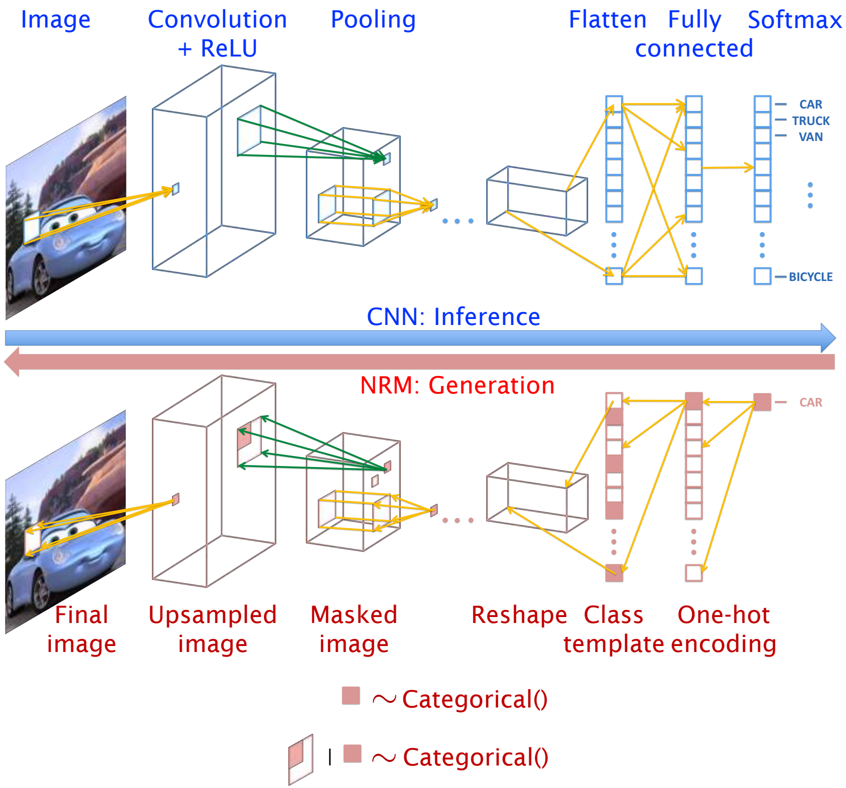
Correspondence between NRM and CNN



Correspondence between NRM and CNN



Losses for Semi-Supervised Learning



- Conditional log-likelihood in the NRM ⇒ Cross-entropy loss for training CNNs with labeled data

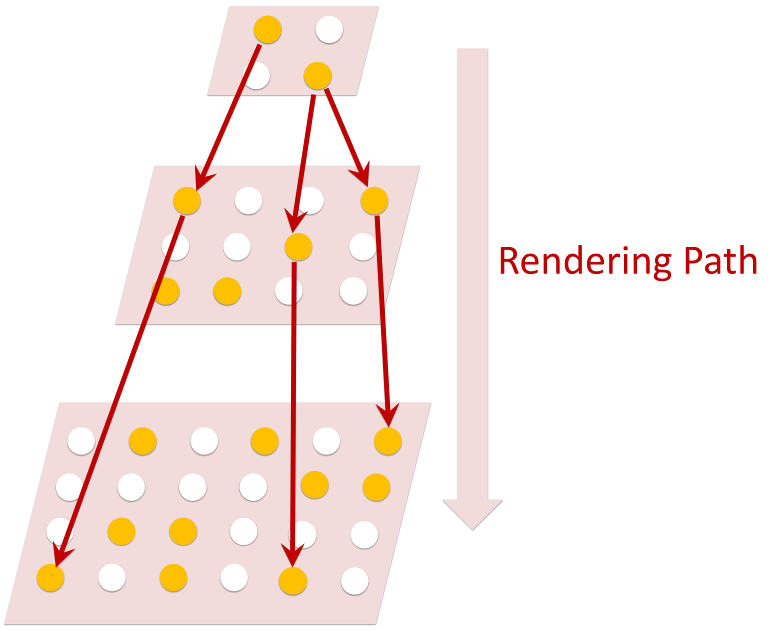
- Expected complete-data log-likelihood in the NRM ⇒ Reconstruction loss + Rendering Path Normalization (RPN) regularizer for training CNNs with unlabeled data

Statistical Guarantees for NRM

NRM is consistent - when training NRM with enough data, the objective value will converge to the optimal one

Bound on the generalization error

$$\text{Risk} \leq \frac{\textit{Number of active rendering paths}}{n^{1/2}}$$



Max-min cross-entropy ➡ max-min networks

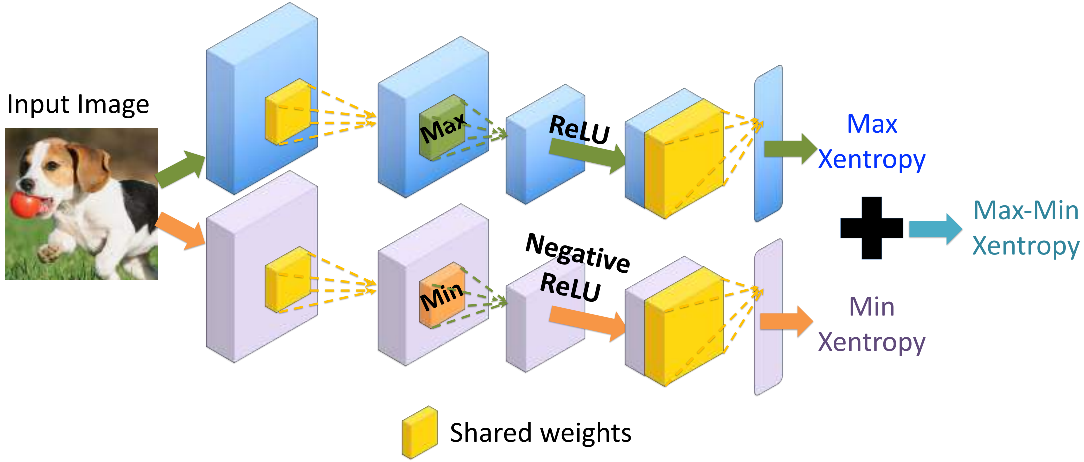
Cross-Entropy Loss for Training the CNNs with Labeled Data

$$\min_{\theta \in \mathcal{A}_\gamma} H_{p,q}(y|x, z^{\max})$$

$$\geq \min_{(z_i)_{i=1}^n, \theta} \frac{1}{n} \sum_{i=1}^{n} -\log p(y_i|x_i, z_i; \theta)$$

Max-Min Loss for Training the CNNs with Labeled Data

$$\alpha^{\max} H_{p,q}(y|x, z^{\max}) + \alpha^{\min} H_{p,q}(y|x, z^{\min})$$



- Max-Min cross-entropy maximizes the posteriors of correct labels and minimizes the posteriors of incorrect labels.
- Co-learning: Max and Min networks try to learn from each other

Empirical results

Semi-Supervised Learning

- NRM + Max-Min improves SOTA by 0.7%-1.8% on CIFAR10, CIFAR100, SVHN.
- Especially in low labeled data setting.

Supervised Learning

- Max-Min improves SOTA on CIFAR10 by 0.26% and on ImageNet by 0.17% (top 5 error)

Max-Min NRM achieves competitive/state-of-the-art results on semi-supervised & supervised learning

RICE

Semi-Supervised Learning Results

Error rate percentage on CIFAR-100

	10K labels 50K images	50K labels 50K images
Π model (Laine & Aila, 2017)	39.19 ± 0.36	26.32 ± 0.04
Temporal Ensembling (Laine & Aila, 2017)	38.65 ± 0.51	26.30 ± 0.15
Supervised-only	44.56 ± 0.30	26.42 ± 0.17
NRM+RPN+Max-Min	$\mathbf{37.75 \pm 0.66}$	$\mathbf{24.38 \pm 0.29}$

Error rate percentage on CIFAR-10

	1K labels 50K images	50K labels 50K images
Mean Teacher (Tarvainen & Valpola, 2017)	21.55 ± 1.48	5.94 ± 0.15
Supervised-only	46.43 ± 1.21	5.82 ± 0.15
NRM+RPN+Mean Teacher+Max-Min	$\mathbf{19.79 \pm 0.74}$	$\mathbf{4.88 \pm 0.09}$

Supervised Learning Results

Error rate percentage on CIFAR-100 and ImageNet

	CIFAR10	ImageNet (top 5 error)
Baseline	2.56	7.21
Max-Min	$\mathbf{2.30 \pm 0.02}$	$\mathbf{7.04}$

Conclusions

- We develop the Neural Rendering Model (NRM) whose bottom-up inference corresponds to a CNN architecture of choice
- We derive consistency guarantee and generalization bound for the NRM and CNNs
- We propose the new Max-Min network and the new Max-Min cross-entropy loss function for learning
- Our Max-Min network achieve state-of-the-art empirical results for semi-supervised and supervised learning

RICE

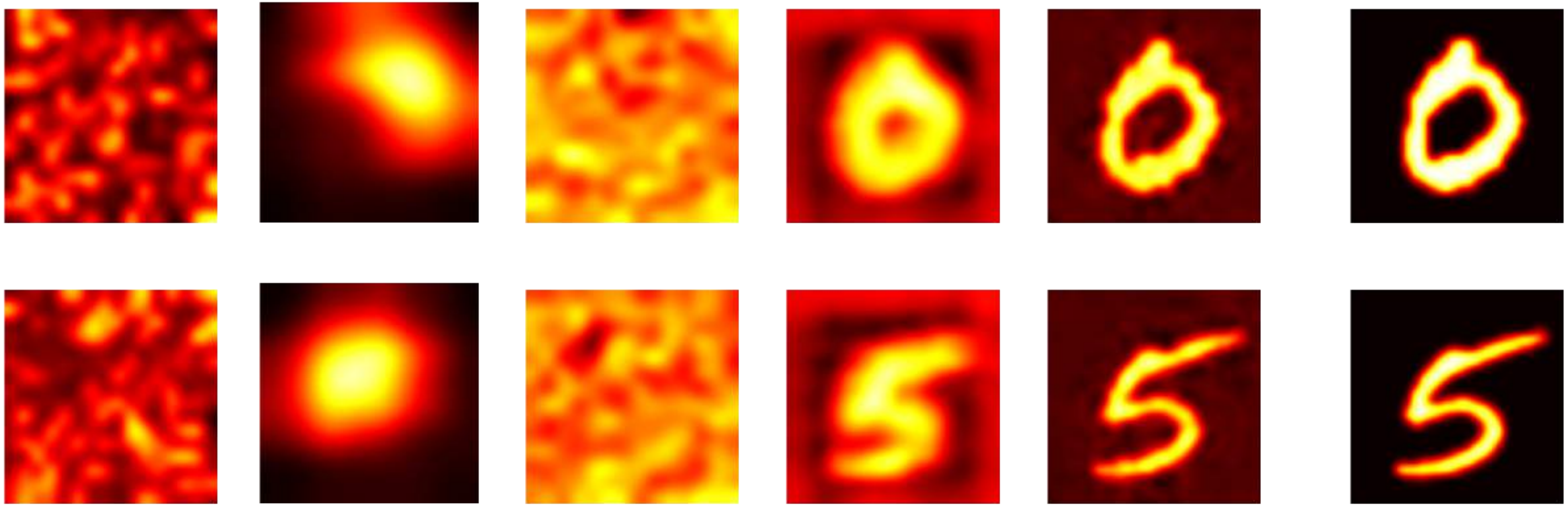
Future Work

- Develop the dynamic NRM to train on time-series data
- Combine NRM and GANs to generate better-looking images
- Use the likelihood from the NRM for out-of-distribution detection
- Explore different noise distributions to derive new CNNs

Back-up Slides

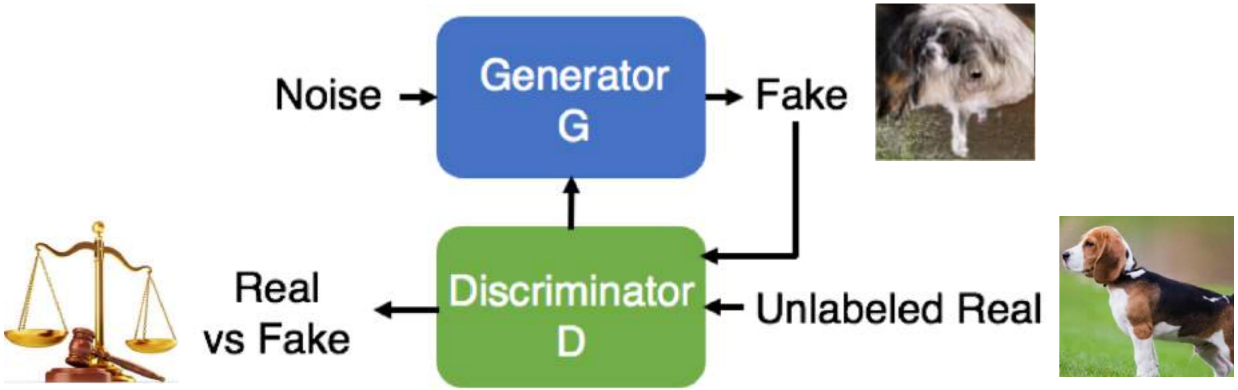
Reconstructed Images from NRM

Layer 4 Layer 3 Layer 2 Layer 1 Layer 0 Original



Semi-Supervised Learning with Generative Models?

Generative Adversarial Networks



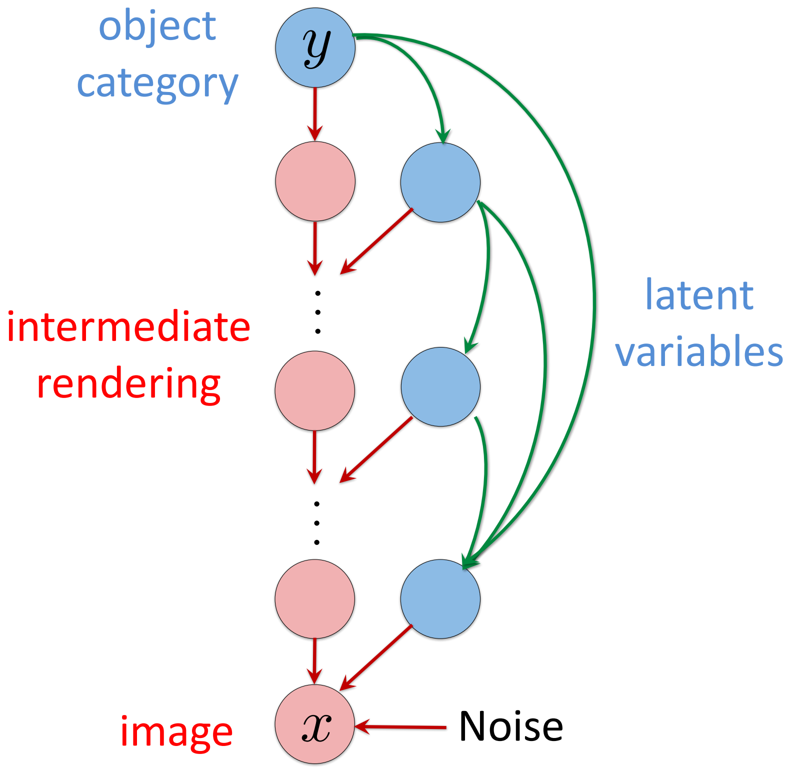
Peril

- Feedback is real vs. fake: different from prediction
- Introduces artifacts

Merits

- Somehow captures statistics of natural images
- Learnable

Connection to Predictive Coding



- The latent dependency in NRM ≈ **backward connections** in predictive coding
- The joint prior of NRM adds bias terms after convolutions in CNNs ≈ backward/feedback neurons in the backward pathways in the brain