CS 4803 / 7643: Deep Learning

Topics:

- (Brief) discussion of multi-modal work
- Fairness, bias
- Wrap-up

Zsolt Kira

Georgia Tech

Lots of Data Modalities

Images

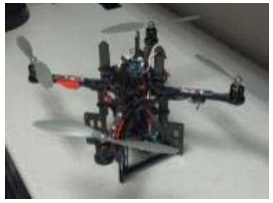
Videos

Speech/Audio

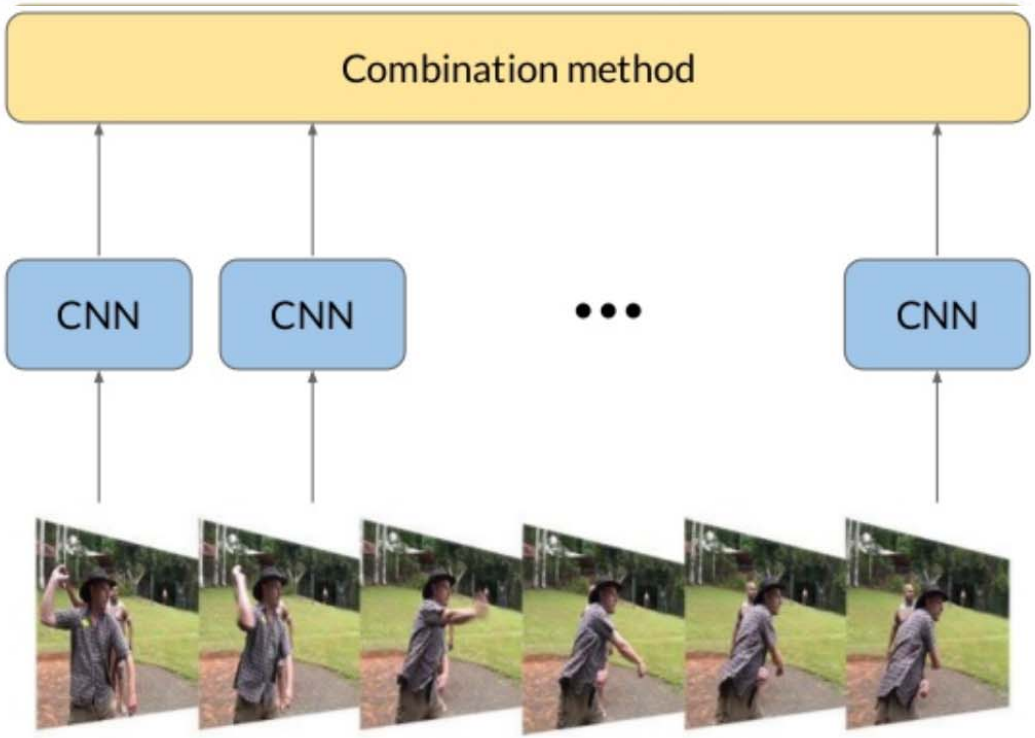
Text

Combined

Sensors



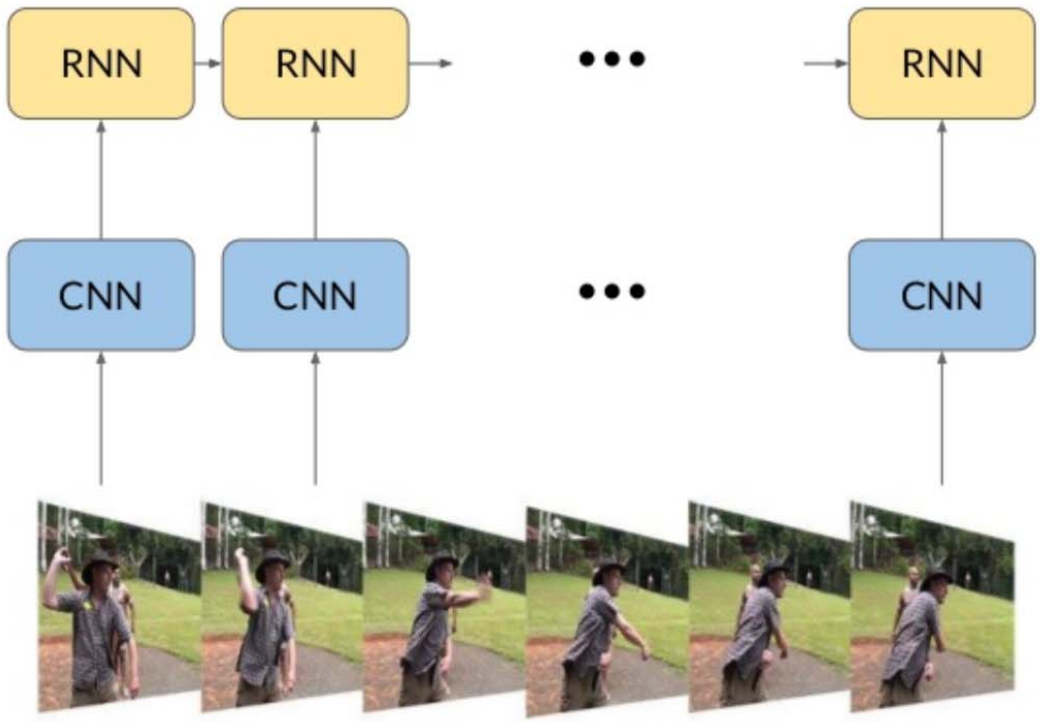
How should we combine them?



Combination is commonly implemented as a small NN on top of a pooling operation (e.g. max, sum, average).

Drawback: pooling is not aware of the temporal order!

Ng et al., Beyond short snippets: Deep networks for video classification, CVPR 2015

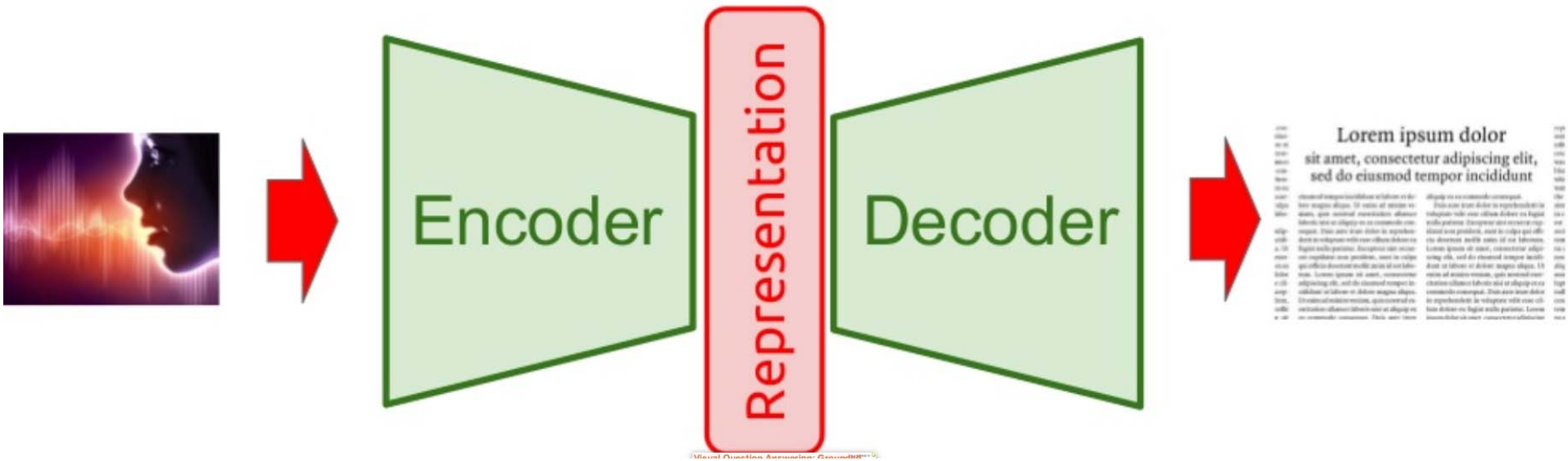


Recurrent Neural Networks are well suited for processing sequences.

Drawback: RNNs are sequential and cannot be parallelized.

Donahue et al., Long-term Recurrent Convolutional Networks for Visual Recognition and Description, CVPR 2015

Cross-Modal Transformation

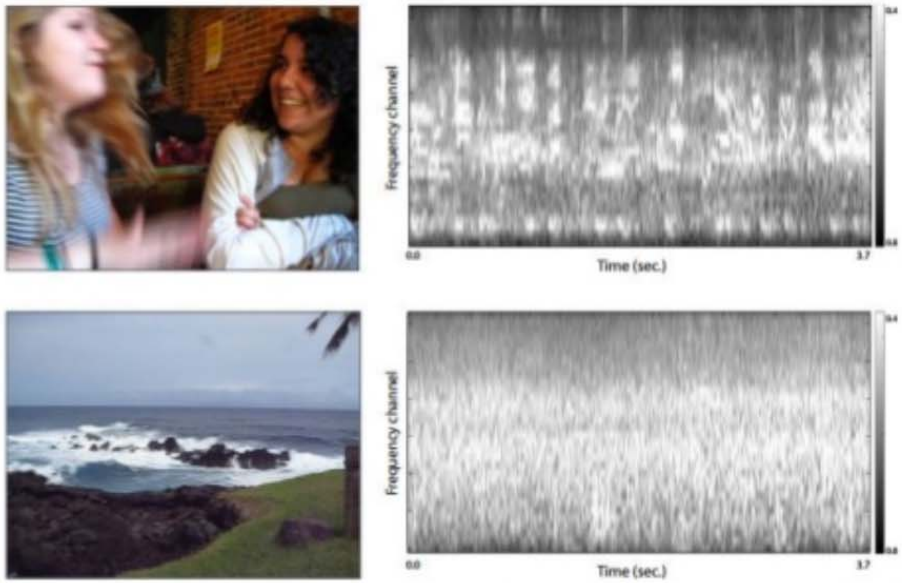


Cross-Modal Transformation

- Lots of examples (some that we've seen)
  - Captioning **(image/video to text)**
  - Text to speech **(text to audio)**
  - Speech to text **(audio to text)**
  - Sign language translation **(video to text)**

Cross-Modal Learning (Unsupervised)

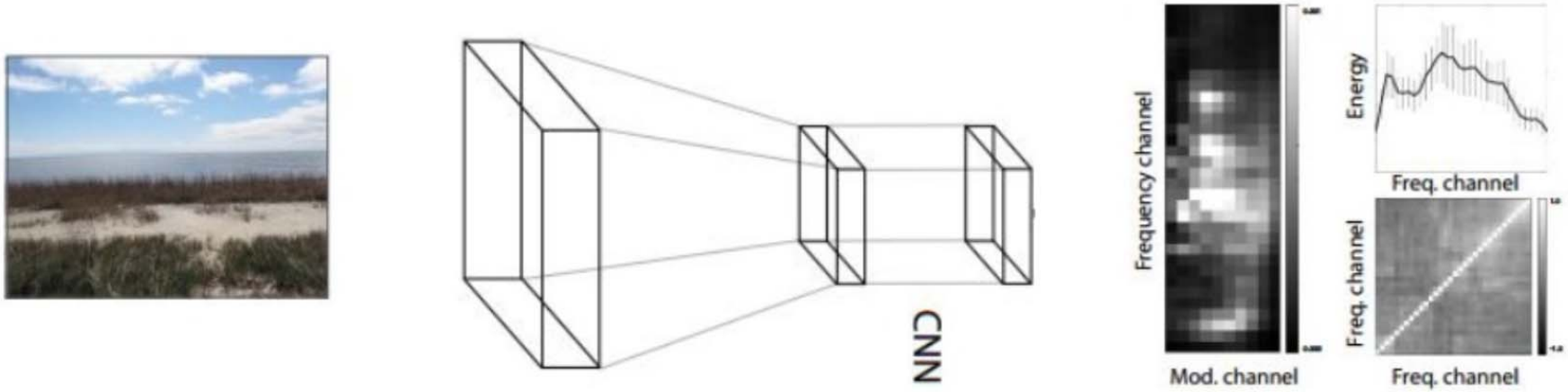
Based on the assumption that ambient sound in video is related to the visual semantics.



(a) Video frame (b) Cochleagram

Owens, Andrew, Jiajun Wu, Josh H. McDermott, William T. Freeman, and Antonio Torralba. "Ambient sound provides supervision for visual learning." ECCV 2016

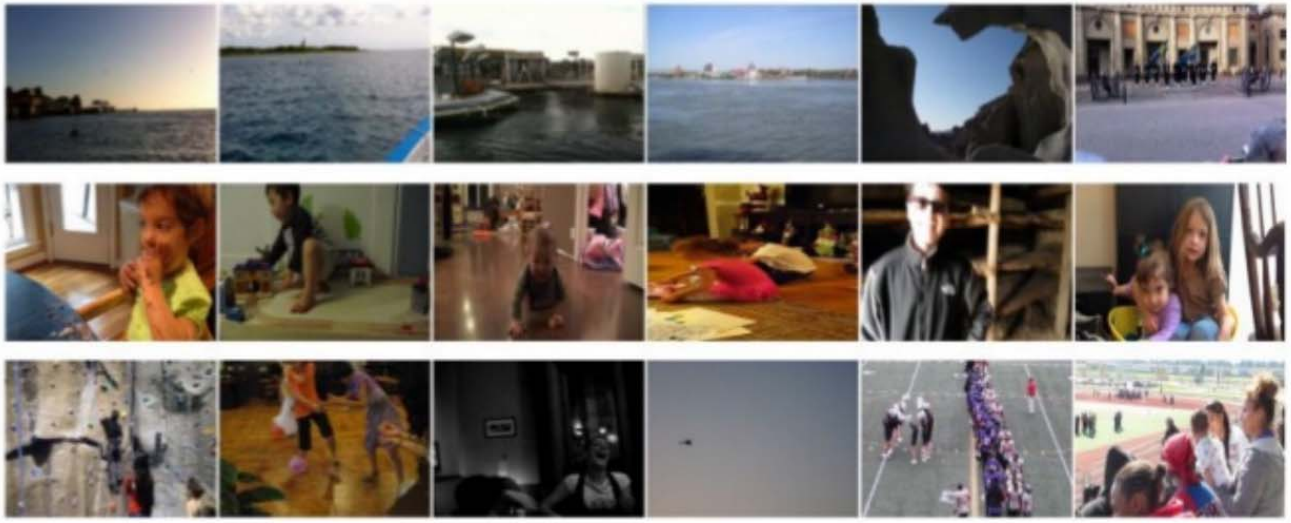
Use videos to train a CNN that predicts the audio statistics of a frame.



Owens, Andrew, Jiajun Wu, Josh H. McDermott, William T. Freeman, and Antonio Torralba. "Ambient sound provides supervision for visual learning." ECCV 2016

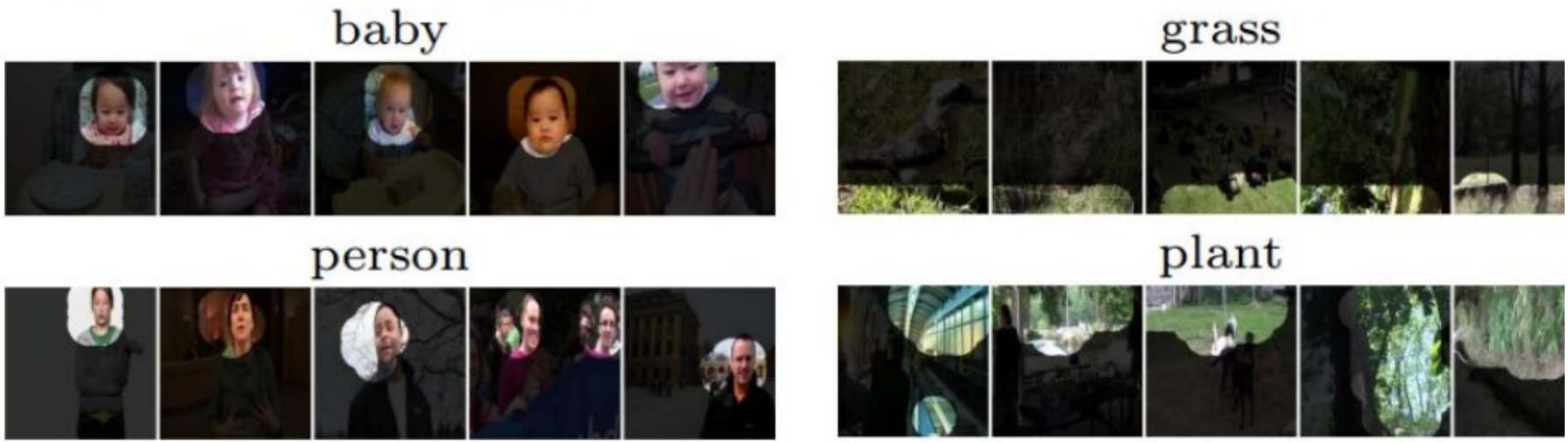
Task: Use the predicted audio stats to clusters images. Audio clusters built with K-means overthe training set

Cluster assignments at test time (one row=one cluster)



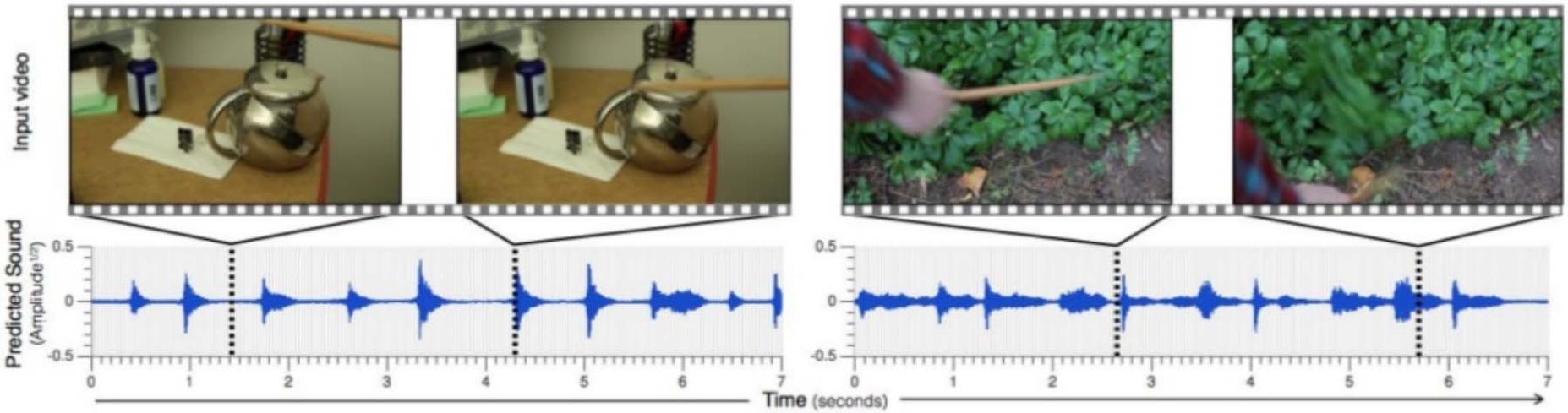
Owens, Andrew, Jiajun Wu, Josh H. McDermott, William T. Freeman, and Antonio Torralba. "Ambient sound provides supervision for visual learning." ECCV 2016

Although the CNN was not trained with class labels, local units with semantic meaning emerge.



Owens, Andrew, Jiajun Wu, Josh H. McDermott, William T. Freeman, and Antonio Torralba. "Ambient sound provides supervision for visual learning." ECCV 2016

Retrieve matching sounds for videos of people hitting objects with a drumstick.

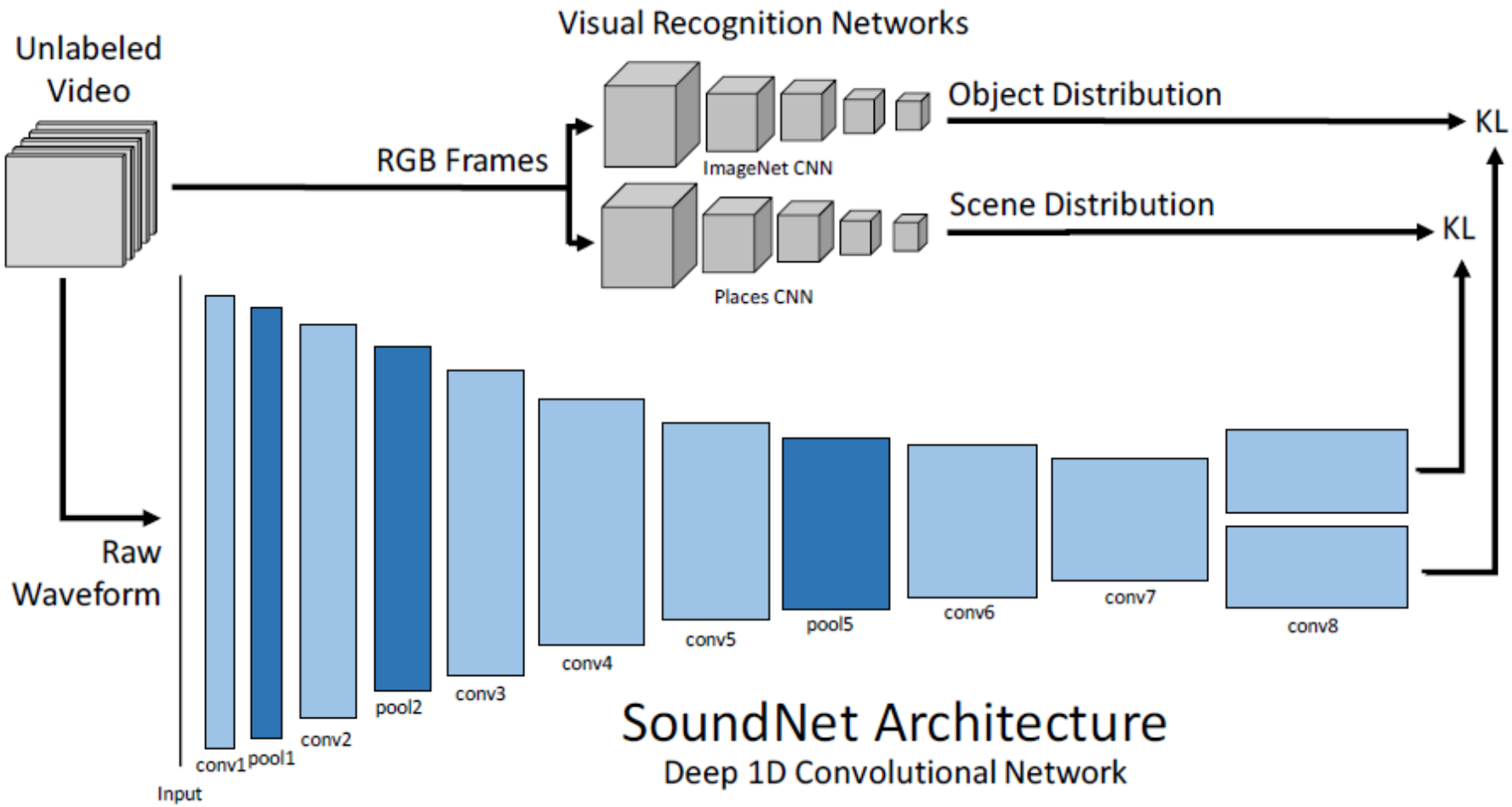


Owens, Andrew, Phillip Isola, Josh McDermott, Antonio Torralba, Edward H. Adelson, and William T. Freeman. "Visually indicated sounds." CVPR 2016.

The Greatest Hits Dataset

Owens, Andrew, Phillip Isola, Josh McDermott, Antonio Torralba, Edward H. Adelson, and William T. Freeman. "Visually indicated sounds." CVPR 2016.

Cross-Modal (Transfer)

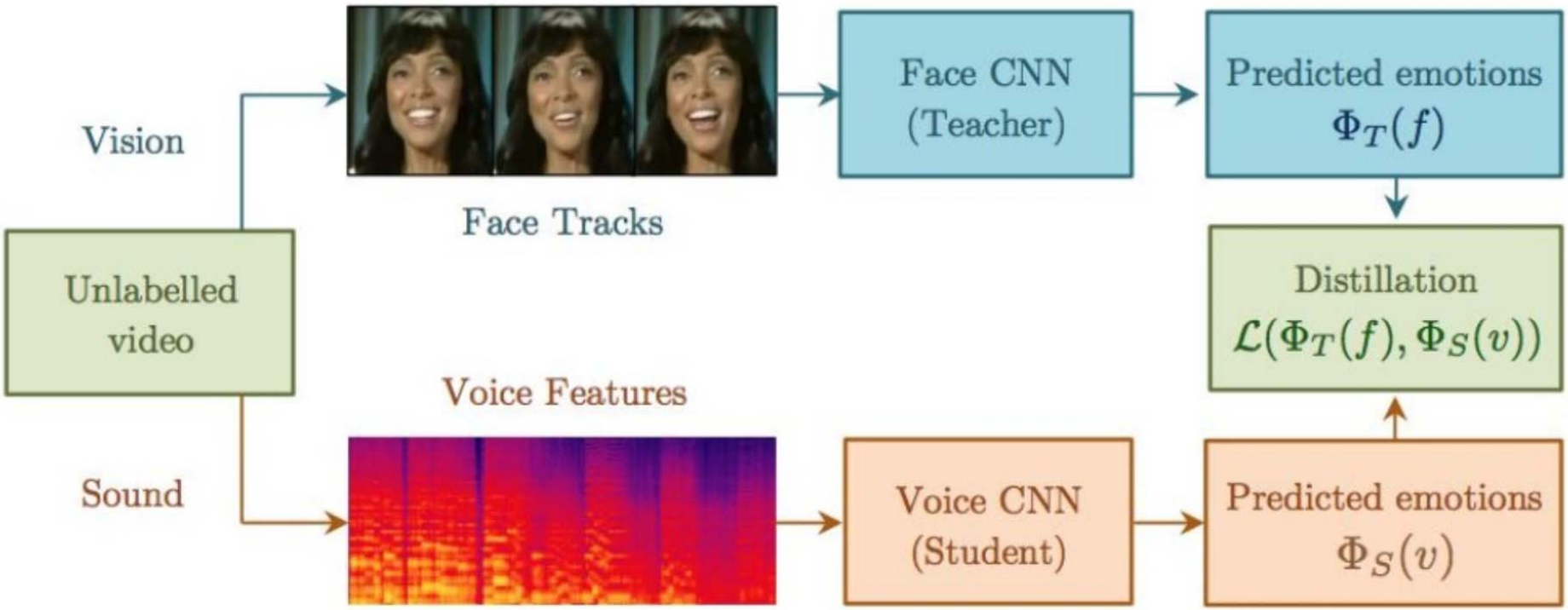


SoundNet: Learning Sound Representations from Unlabeled Video

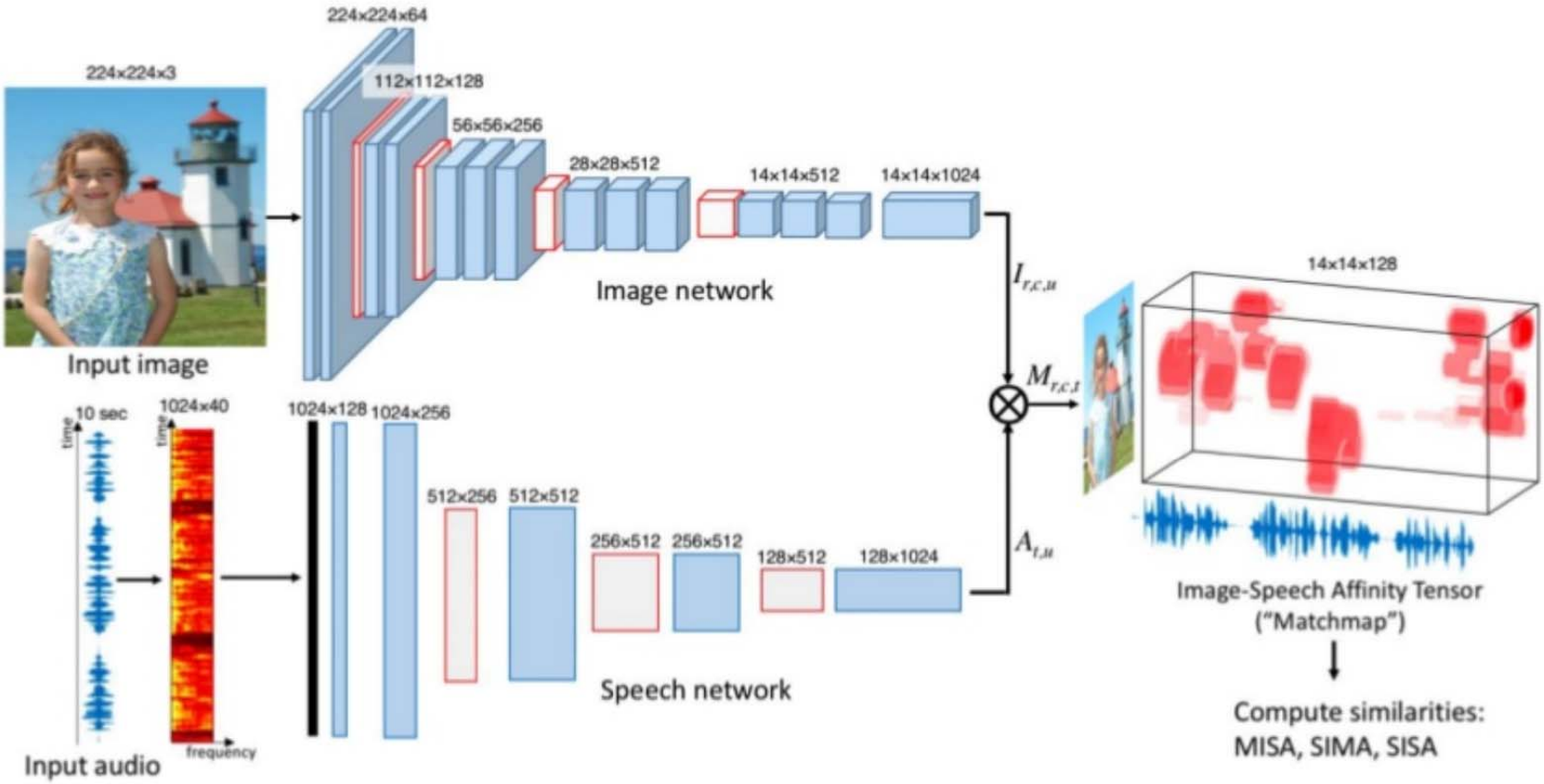
Method	Accuracy
RG [29]	69%
LTT [21]	72%
RNH [30]	77%
Ensemble [34]	78%
SoundNet	**88%**

Method	Accuracy on ESC-50	Accuracy on ESC-10
SVM-MFCC [28]	39.6%	67.5%
Convolutional Autoencoder	39.9%	74.3%
Random Forest [28]	44.3%	72.7%
Piczak ConvNet [27]	64.5%	81.0%
SoundNet	**74.2%**	**92.2%**
Human Performance [28]	81.3%	95.7%

Teacher network: Facial Emotion Recognition (visual)



S Albanie, A Nagrani, A Vedaldi, A Zisserman, "Emotion Recognition in Speech using Cross-Modal Transfer in the Wild"
ACM Multimedia 2018.



Harwath, David, Adrià Recasens, Dídac Surís, Galen Chuang, Antonio Torralba, and James Glass. "Jointly Discovering Visual Objects and Spoken Words from Raw Sensory Input." ECCV 2018.

Mult-Modal Fusion

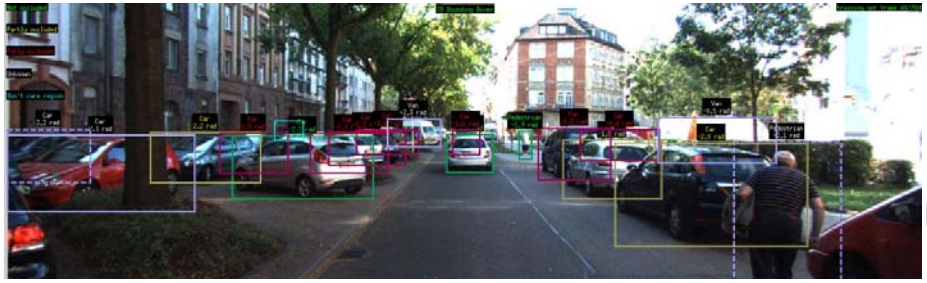
- Several approaches:
  - Just combine features at some level of representation (which?)
  - Gating mechanisms
  - Learn joint feature space

Mult-Modal Fusion

- Several approaches:
  - **Just combine features at some level of representation (which?)**
  - Gating mechanisms
  - Learn joint feature space

The KITTI Dataset

- **Purpose**:
  - To develop algorithms simulating ground-level sensors that will be on quadrotor
  - To be able to publish algorithms (requires comparison to state of the art on a public dataset)
  - Already ground-truthed!



stereo | flow | odometry | object | tracking | road | semantics | raw data

Dense 3D from Sparse LIDAR

From sparse lidar

To a dense depth map

$$D_p = \frac{1}{W_p} \sum_{q \in \mathcal{N}} G_{\sigma_s}(||p-q||) G_{\sigma_r}(|I_q|) I_q \qquad (1)$$

where G_{σ_s} weights points q inversely to their distance to position p, G_{σ_r} penalizes the influence of points as function of their range values, and finally W_p is a normalization factor that ensures weights sum to one, *i.e.*, $W_p = \sum_{q \in \mathcal{N}} G_{\sigma_s}(||p-q||) G_{\sigma_r}(|I_q|)$.

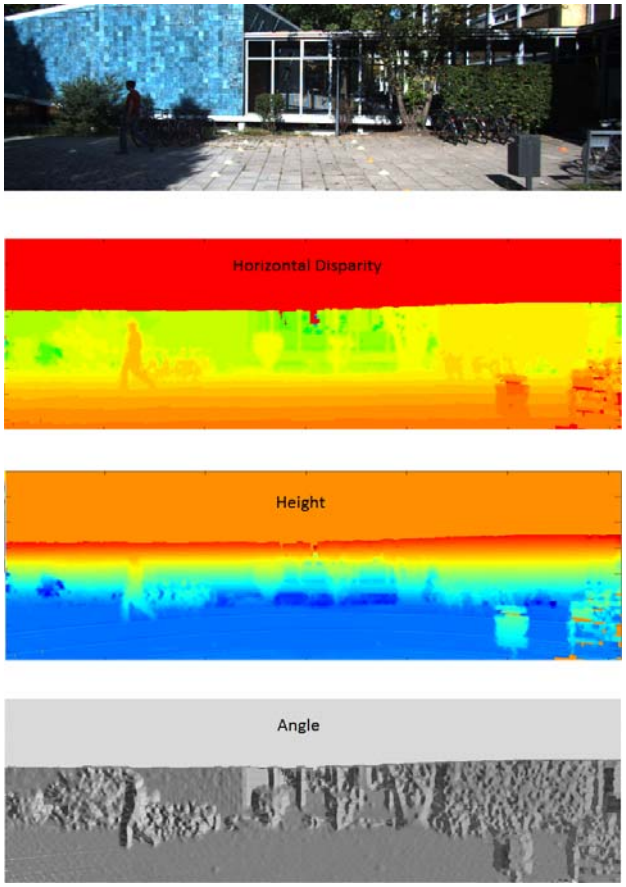
Via spatial filtering

Premebida, C., Carreira, J., Batista, J., Nunes, U., 2014. Pedestrian detection combining RGB and dense LIDAR data, in: Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on. IEEE, pp. 4112–4117.

HHA Representation

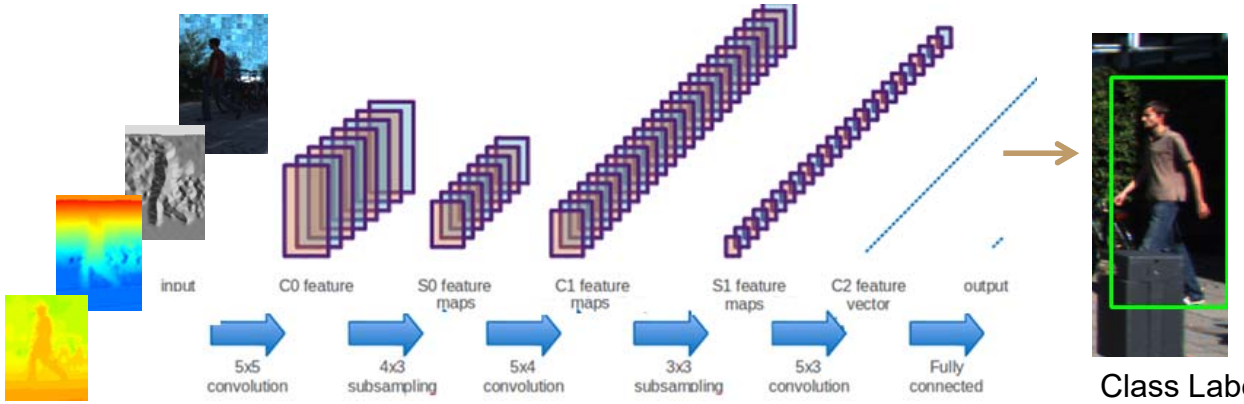
- We have modified algorithm/code to extract **multiple** aspects of 3D information obtained from velodyne LIDAR
- Currently state of art in indoor RGBD tasks
- We have modified it to make it work for unstructured outdoor scenes
- These additional channels of information will be incorporated in next redesigned CNN architecture

Gupta, S., Girshick, R., Arbeláez, P., Malik, J., 2014. Learning rich features from RGB-D images for object detection and segmentation, in: Computer Vision–ECCV 2014. Springer, pp. 345–360.



Object Detection Pipeline

Region Proposals → CNN Feature Extraction → Classification & Box Regression



Class Label,

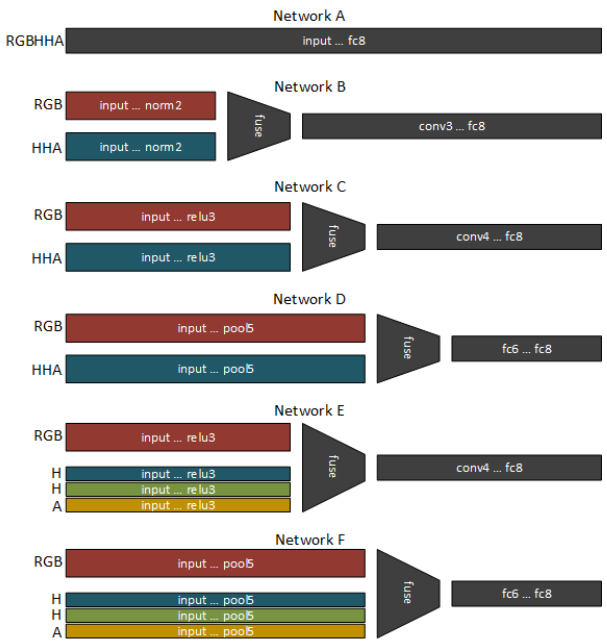
Bounding Box (x,y,w,h)

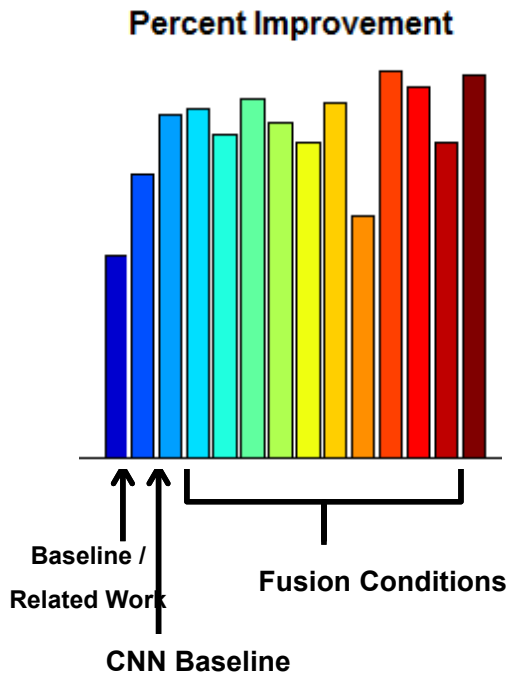
Question: Where should fusion occur?

- Research question: Given that
  - CNNs learn a hierarchical feature space
  - Each channel is representing a different aspect of the scene
- Where should fusion occur?

J. Schlosser, C. Chow, Z. Kira, "Fusing LIDAR and Images for Pedestrian Detection using Convolutional Neural Networks", accepted to the International Conference on Robotics & Automation (ICRA), 2015

Deep Learning Architectures for Multi-Modal Fusion

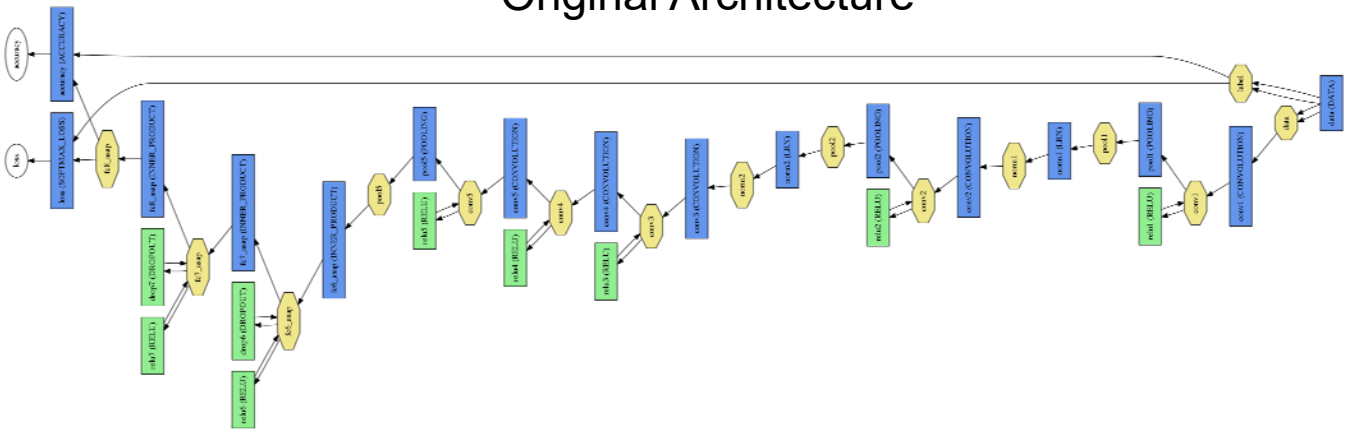




Explore the question: At what level of representation should we fuse different modalities (image and depth)?

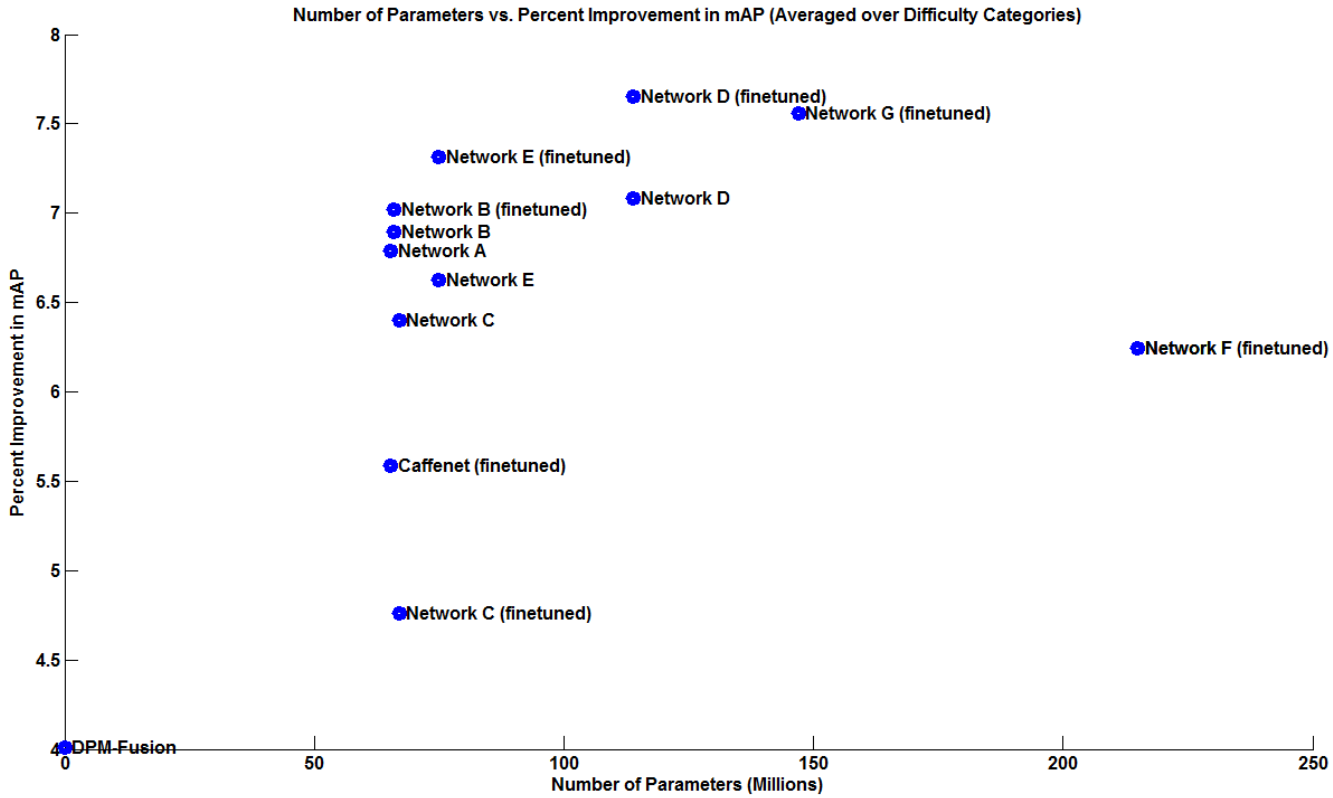
Fusion Architectures

Original Architecture



Fusion Architecture (Fusion-4c-conv3)

Parameter Efficiency vs. Performance



Multi-Modal Fusion

Chung, Joon Son, Andrew Senior, Oriol Vinyals, and Andrew Zisserman. "Lip reading sentences in the wild." *CVPR 2017*

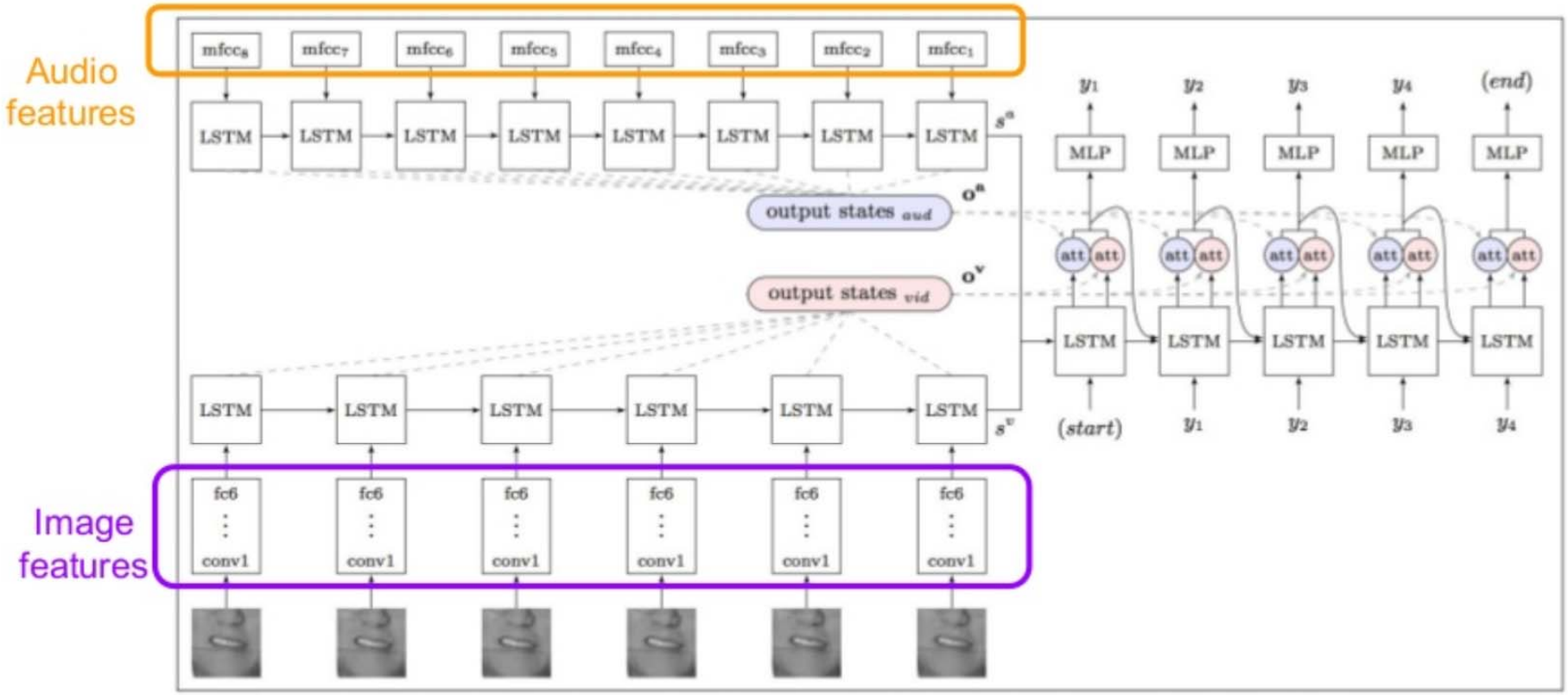


Figure 1. *Watch, Listen, Attend and Spell* architecture. At each time step, the decoder outputs a character y_i, as well as two attention vectors. The attention vectors are used to select the appropriate period of the input visual and audio sequences.

Chung, Joon Son, Andrew Senior, Oriol Vinyals, and Andrew Zisserman. "Lip reading sentences in the wild." *CVPR 2017*

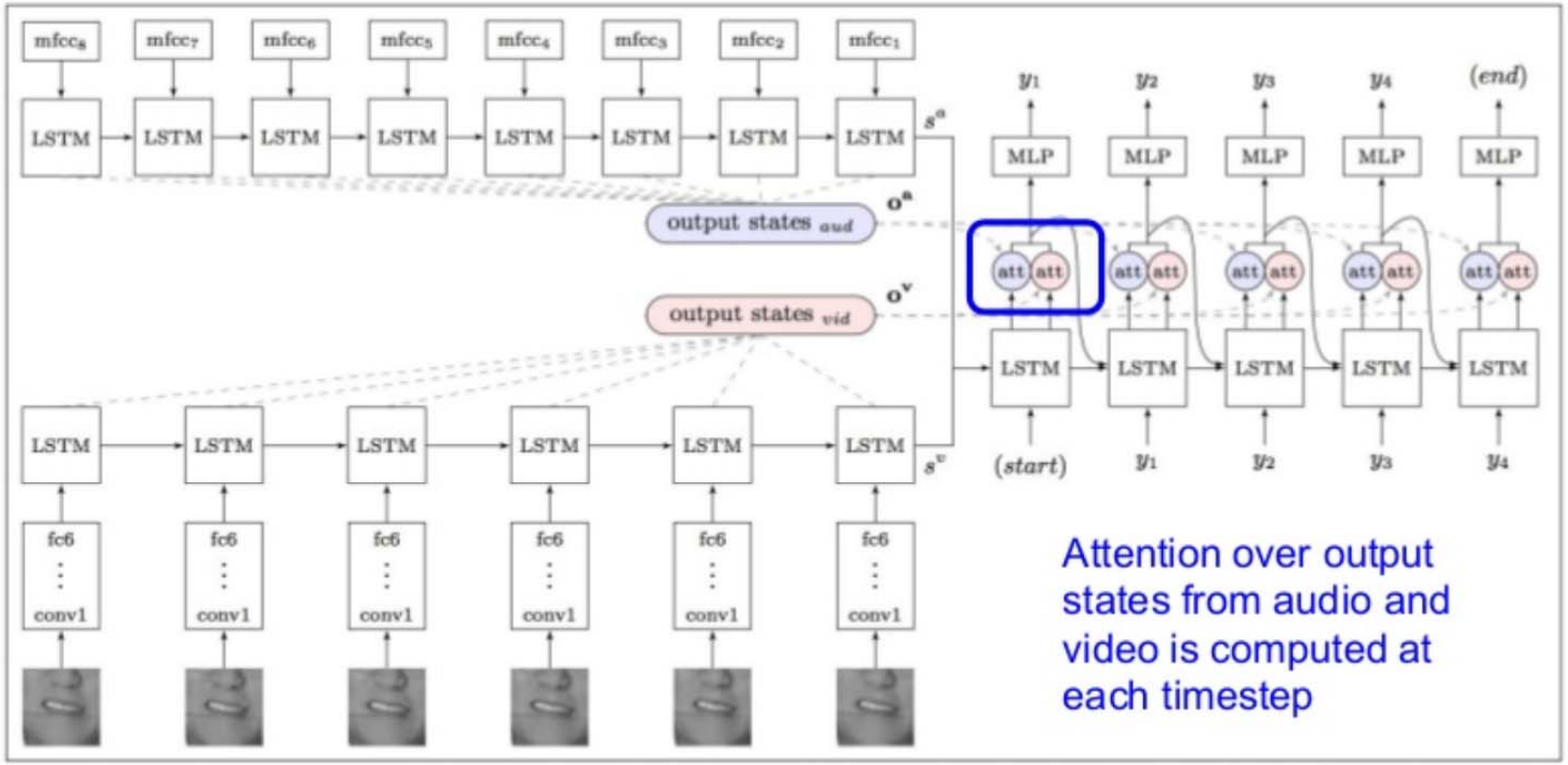


Figure 1. *Watch, Listen, Attend and Spell* architecture. At each time step, the decoder outputs a character y_i, as well as two attention vectors. The attention vectors are used to select the appropriate period of the input visual and audio sequences.

Chung, Joon Son, Andrew Senior, Oriol Vinyals, and Andrew Zisserman. "Lip reading sentences in the wild." *CVPR 2017*

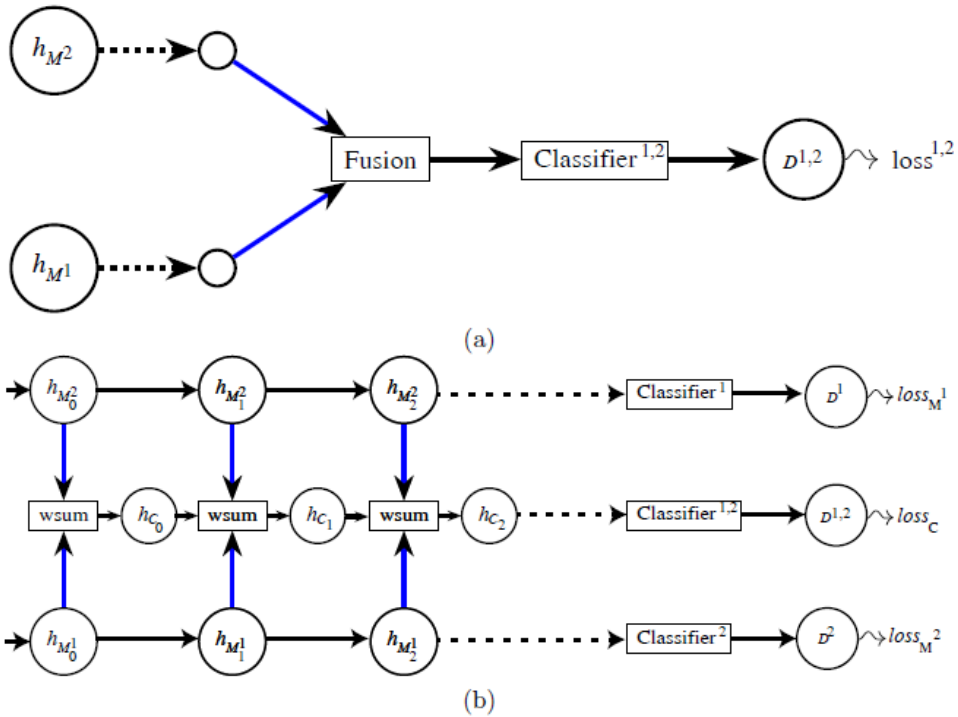


Fig. 2. (a) Basic fusion method, fusing the hidden representations of the modalities at a given layer and then using only joint representation. Fusing at a low-level layer is called early fusion while fusing at the last layer is called late fusion. **(b) Our CentralNet fusion model**, using both unimodal hidden representations and a central joint representation at each layer. The fusion of the unimodal representations is done here using a learned weighted sum. For the sake of simplicity, only the overall synoptic views of the architectures are represented. More details are provided in Section 2.

CentralNet: a Multilayer Approach for Multimodal Fusion, Valentin Vielzeuf, Alexis Lechervy, Stéphane Pateux, Frédéric Jurie

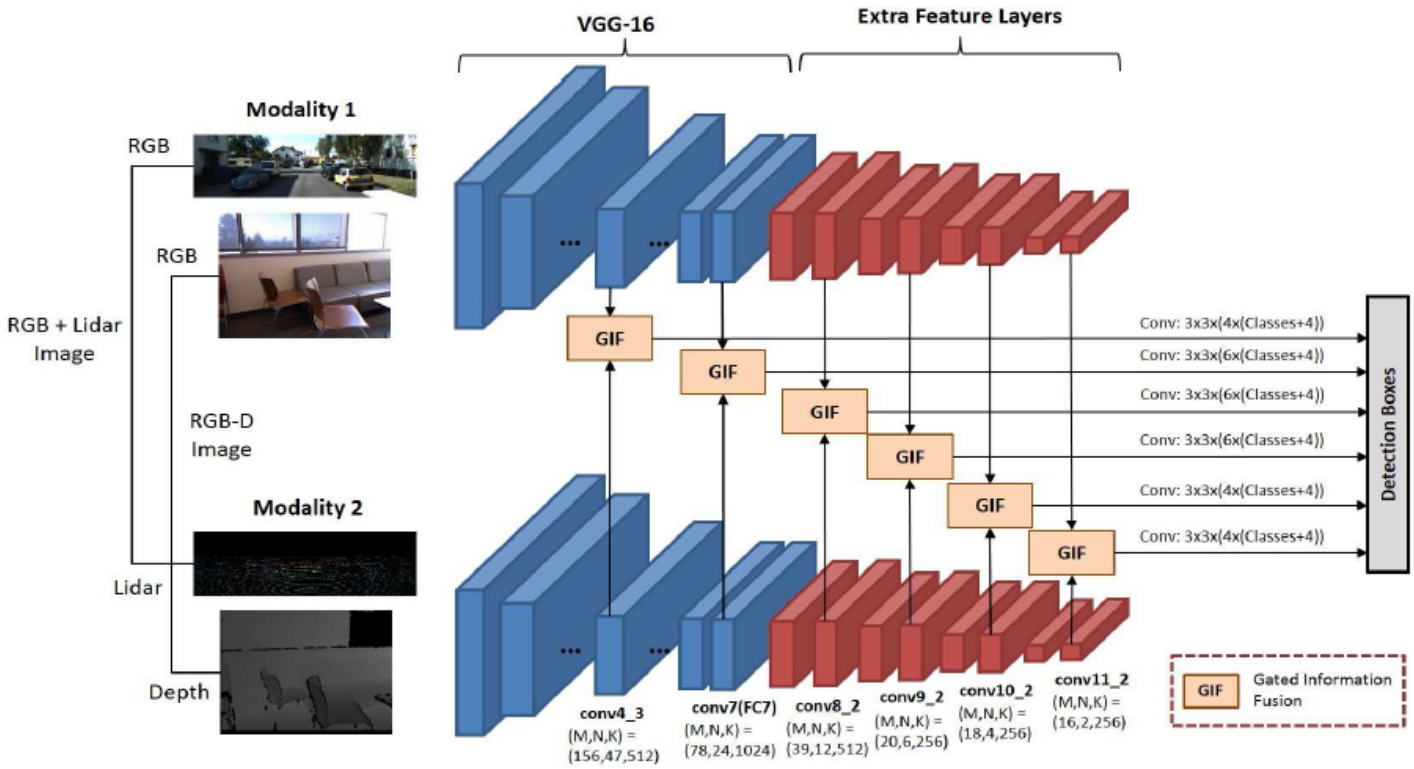


Fig. 1. Overall structure of the proposed R-DML. The R-DML takes the intermediate feature maps from both modality 1 and modality 2 using separate CNNs and combines them through the proposed GIF network. The joint feature maps produced by the GIF network are used to compute the score for object detection following the procedure of SSD.

Robust Deep Multi-modal Learning Based on Gated Information Fusion Network, Kim, et al.

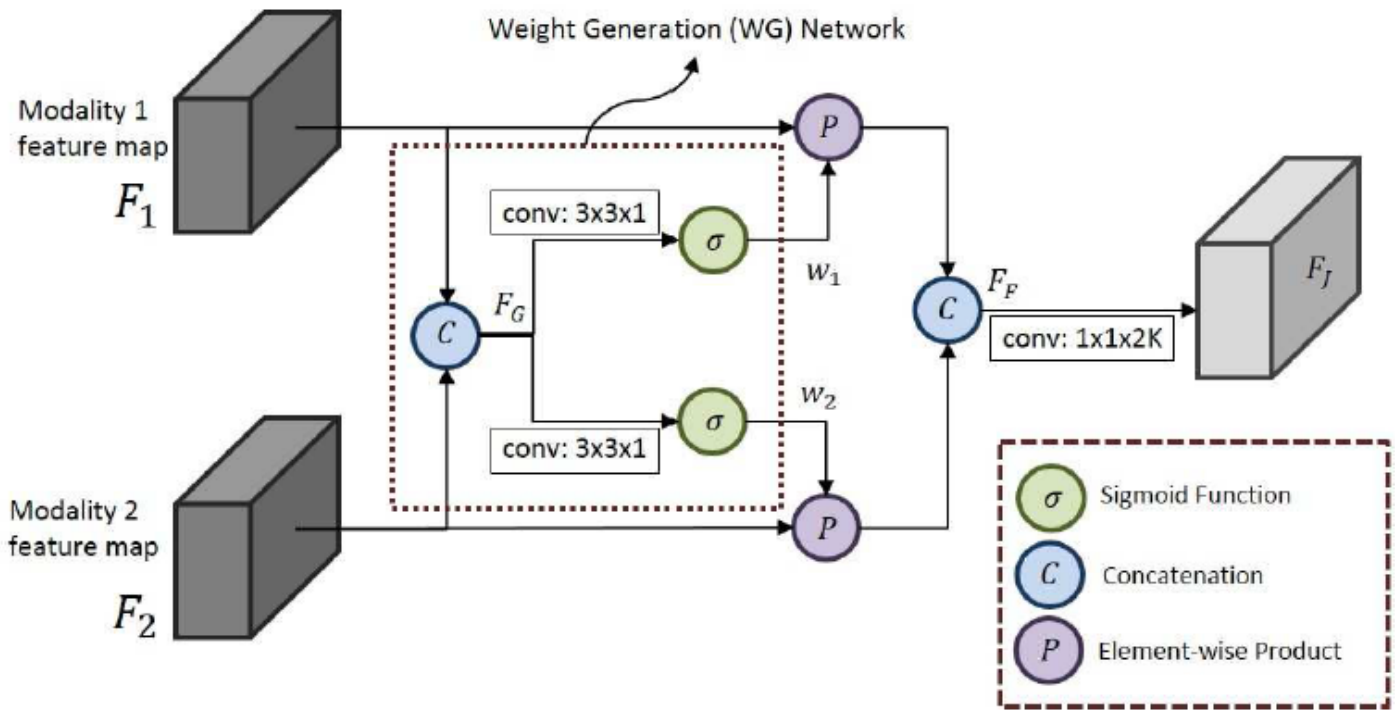
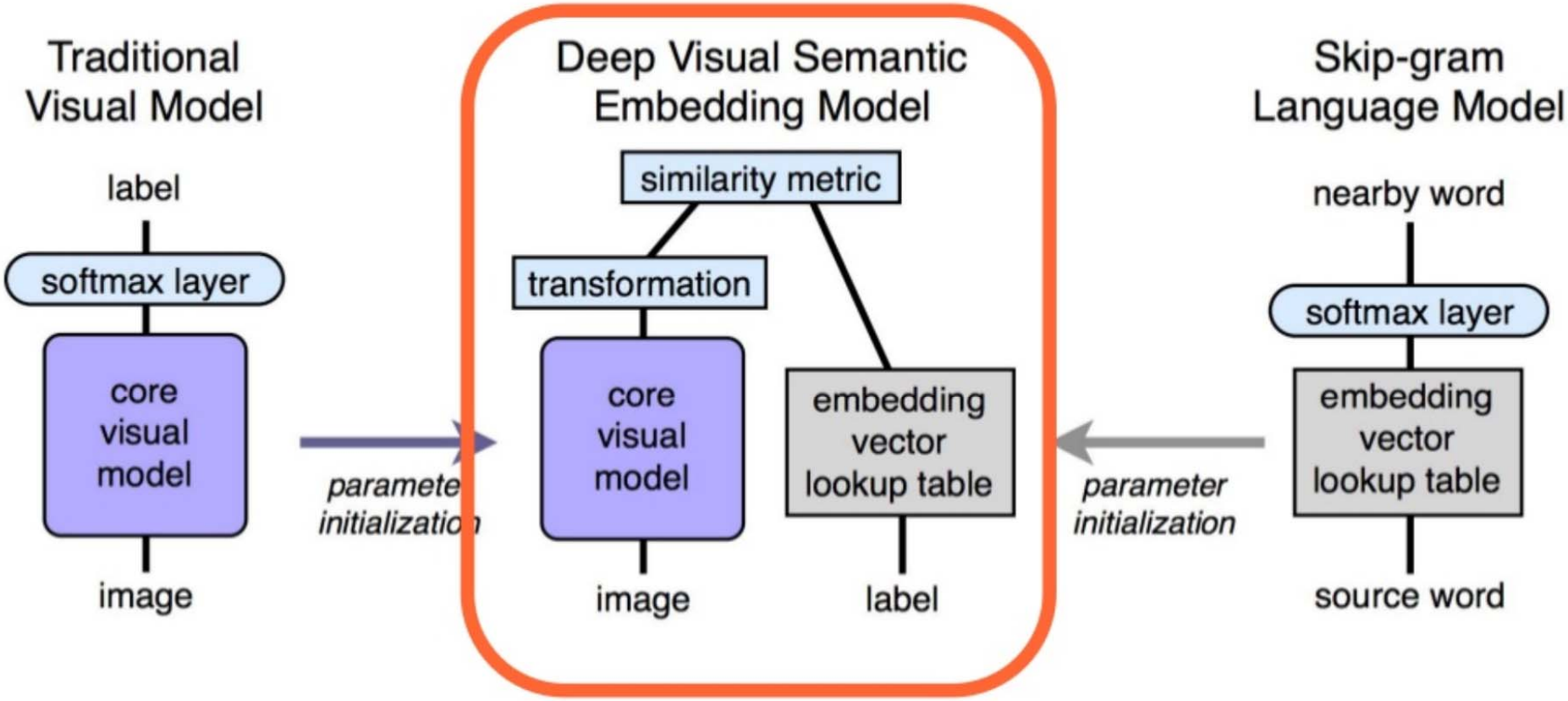


Fig. 2. The structure of the proposed GIF network. The GIF network produces the weight maps $\mathbf{w}_1$ and $\mathbf{w}_2$ by applying the convolutional layer and sigmoid function to the input features. Then, $\mathbf{w}_1$ and $\mathbf{w}_2$ are multiplied to the feature maps $\mathbf{F}_1$ and $\mathbf{F}_2$ for weighted information fusion.

Robust Deep Multi-modal Learning Based on Gated Information Fusion Network, Kim, et al.

Mult-Modal Fusion

- Several approaches:
  - Just combine features at some level of representation (which?)
  - Gating mechanisms
  - **Learn joint feature space**



Frome, Andrea, Greg S. Corrado, Jon Shlens, Samy Bengio, Jeff Dean, and Tomas Mikolov. "Devise: A deep visual-semantic embedding model." NIPS 2013

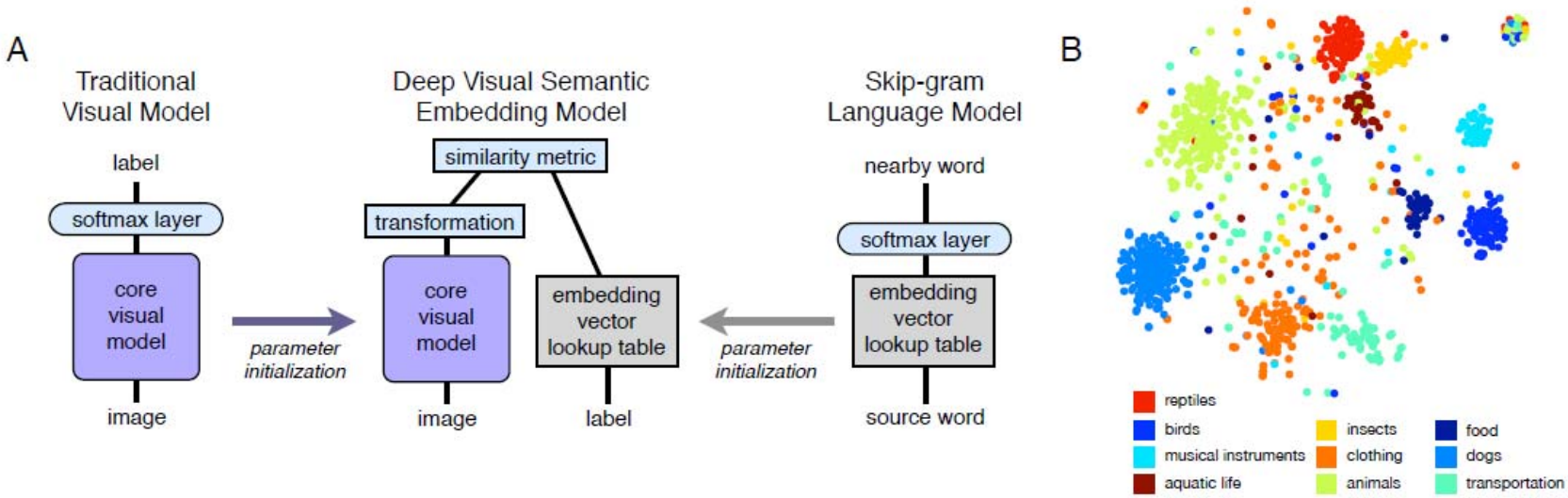
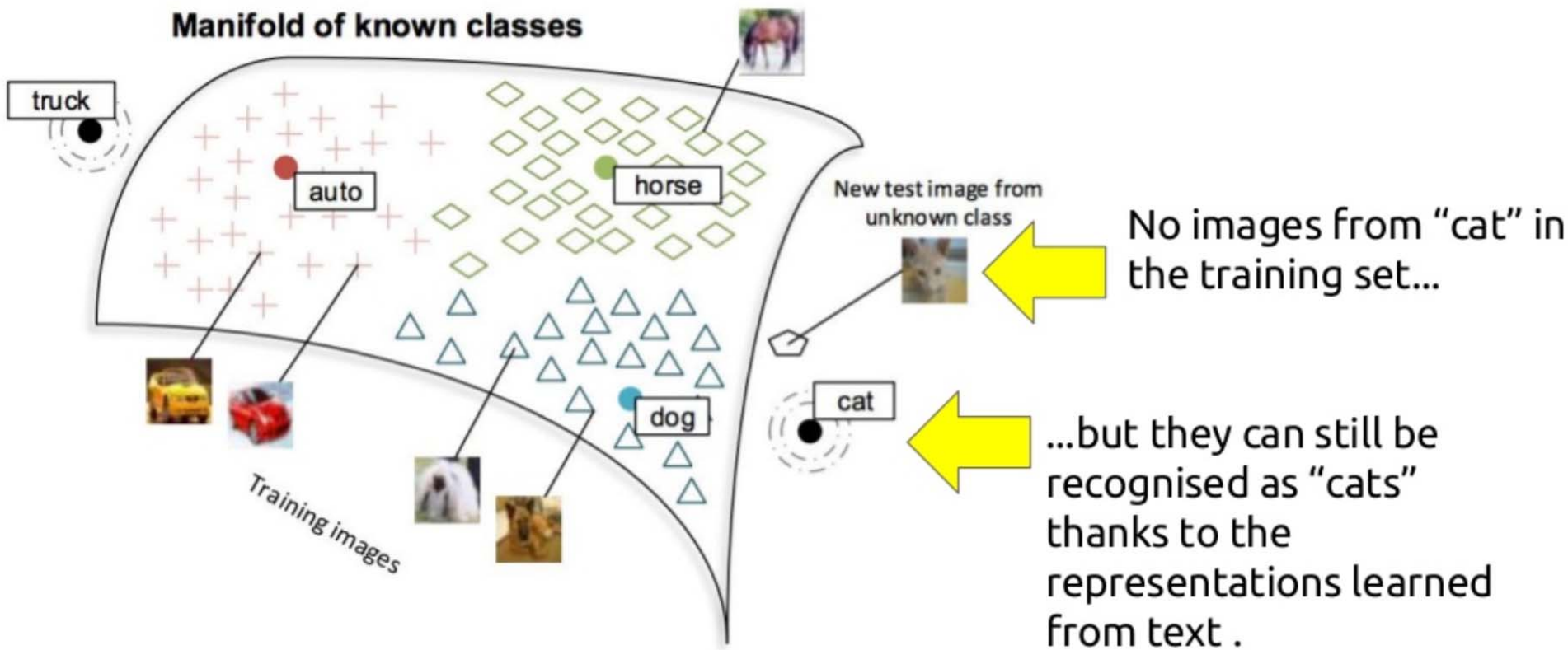
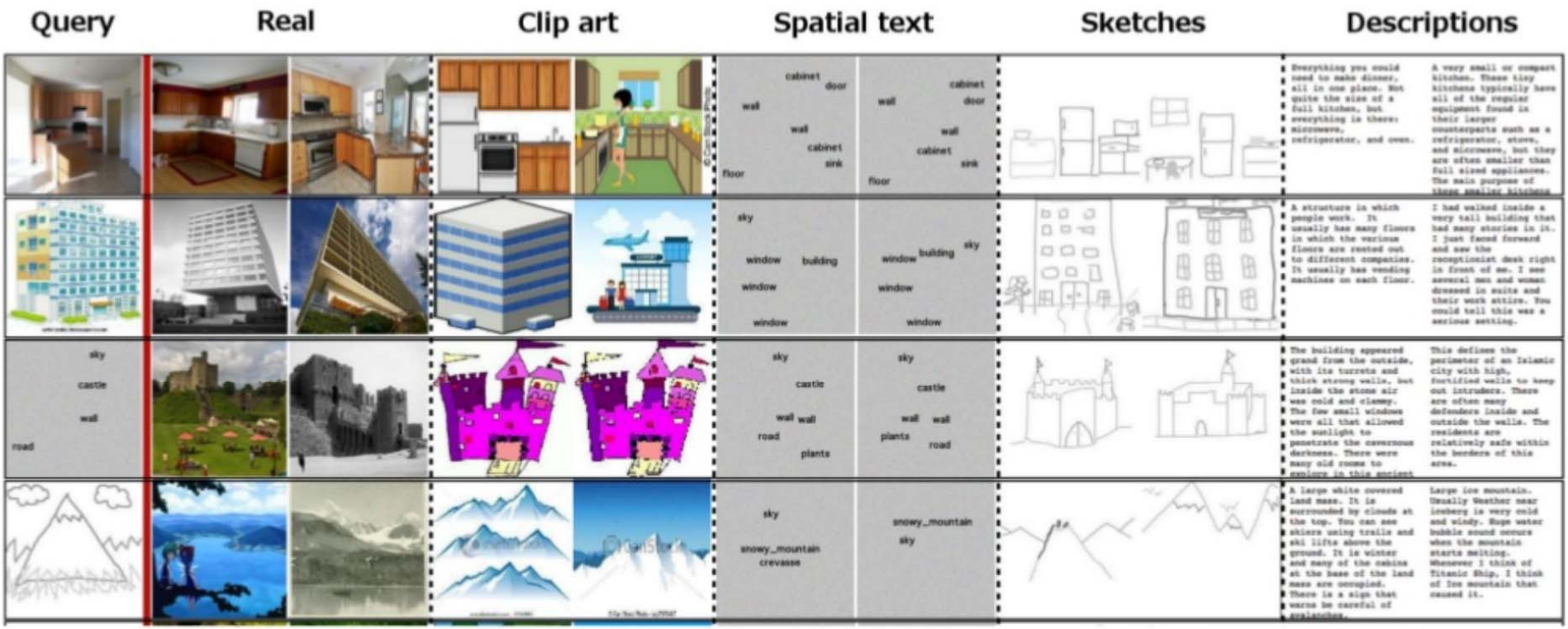


Figure 1: (a) Left: a visual object categorization network with a softmax output layer; Right: a skip-gram language model; Center: our joint model, which is initialized with parameters pre-trained at the lower layers of the other two models. (b) t-SNE visualization [19] of a subset of the ILSVRC 2012 1K label embeddings learned using skip-gram.



Socher, R., Ganjoo, M., Manning, C. D., & Ng, A., Zero-shot learning through cross-modal transfer. NIPS 2013 [slides] [code]



Aytar, Yusuf, Lluis Castrejon, Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. "Cross-Modal Scene Networks." CVPR 2016.

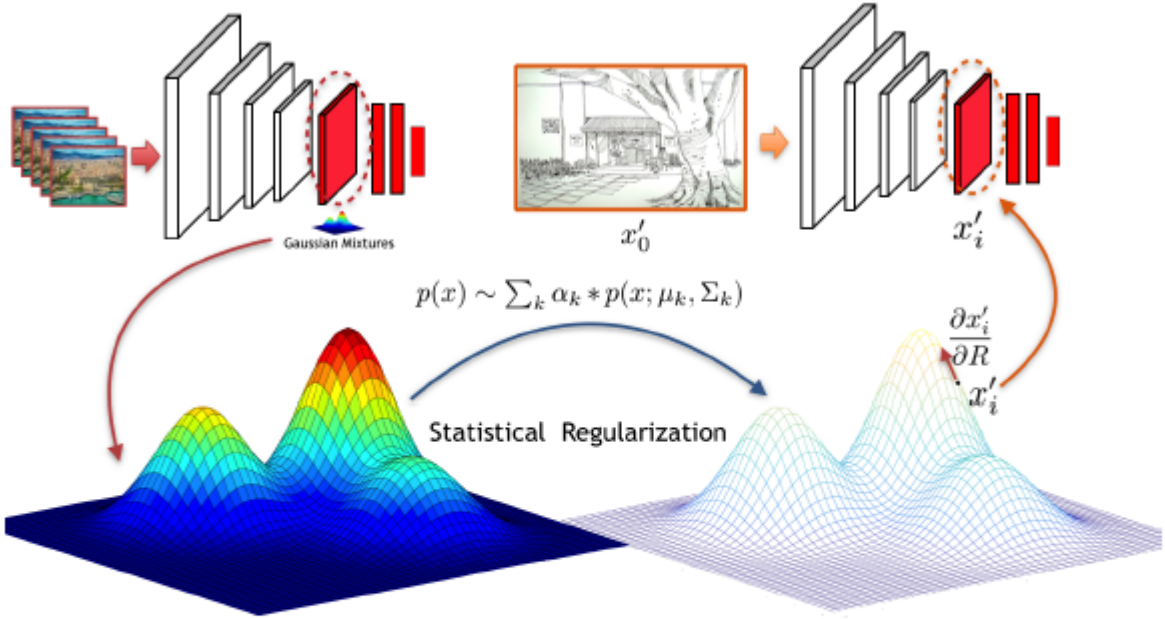


Fig. 4: **Statistical Regularization.** We illustrate this regularization with an example. Above, the feature distribution $p(x_i)$ learned from Places network is modeled with a GMM, and on incorporated as a prior on x_i while optimizing the deep model in line drawings modality.

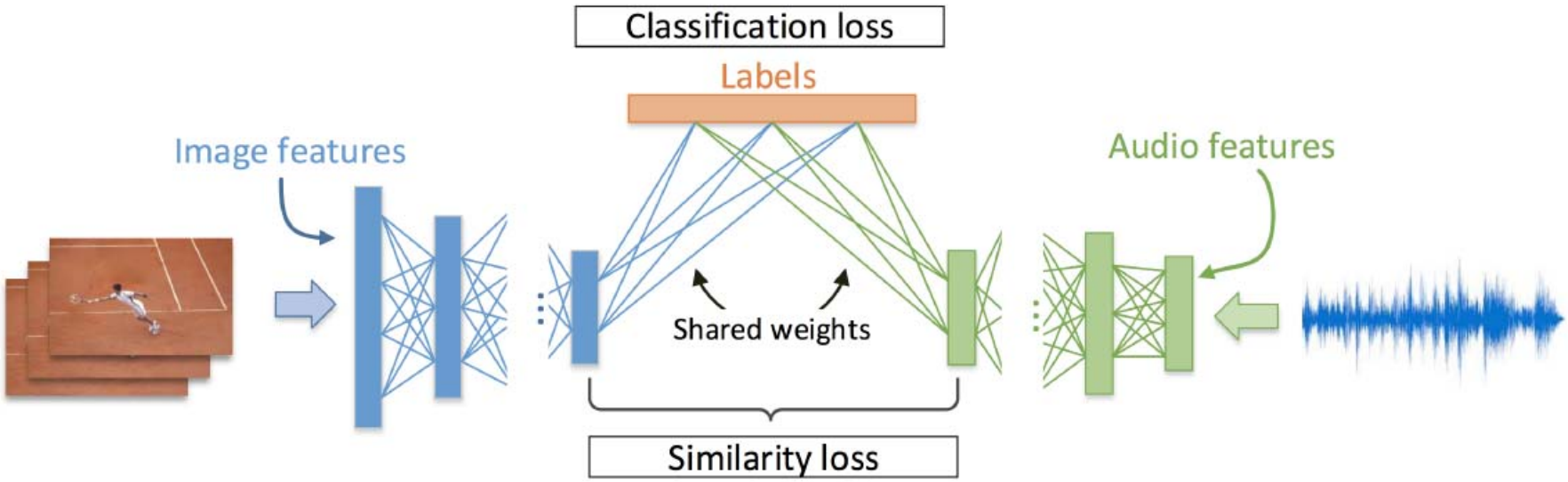
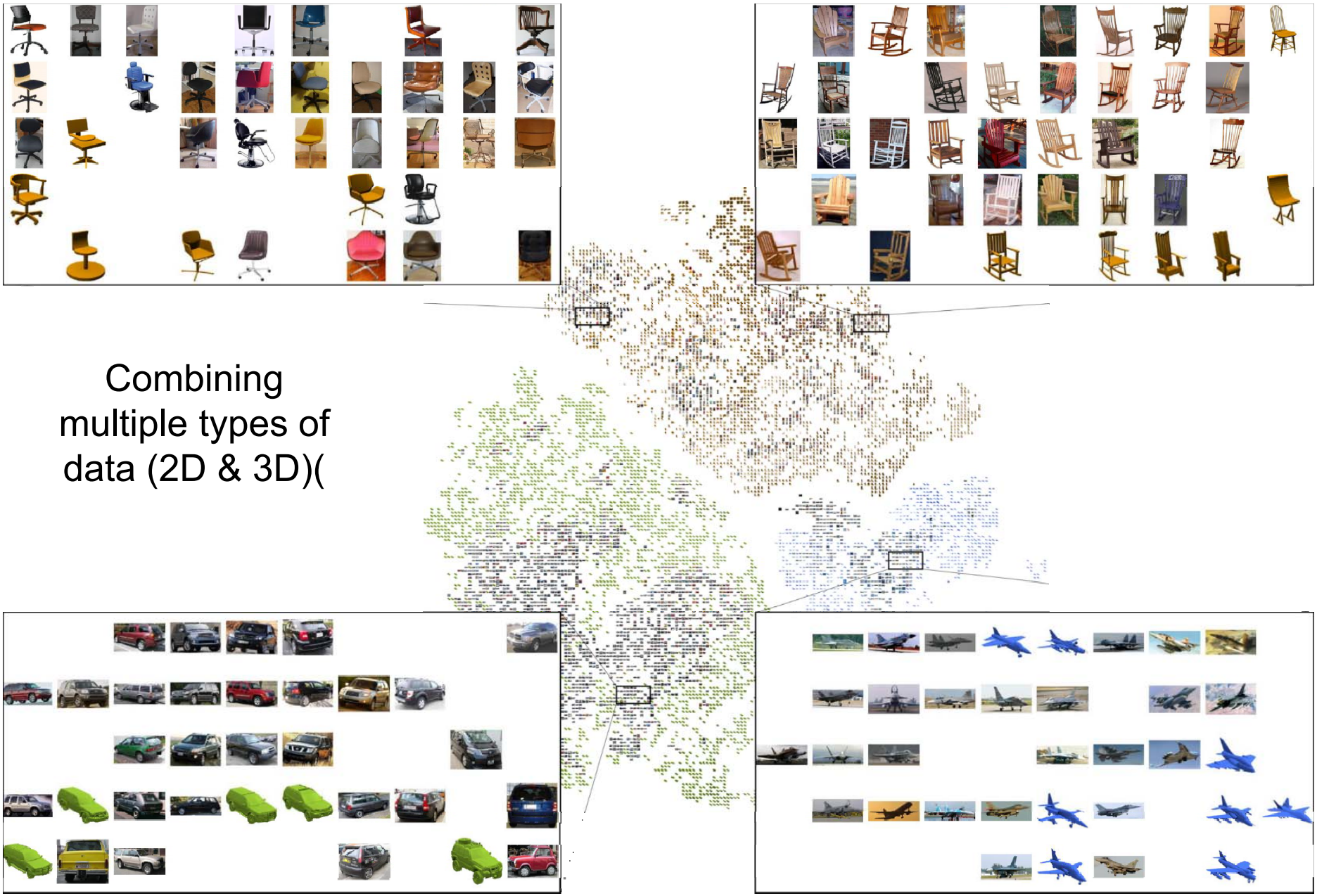


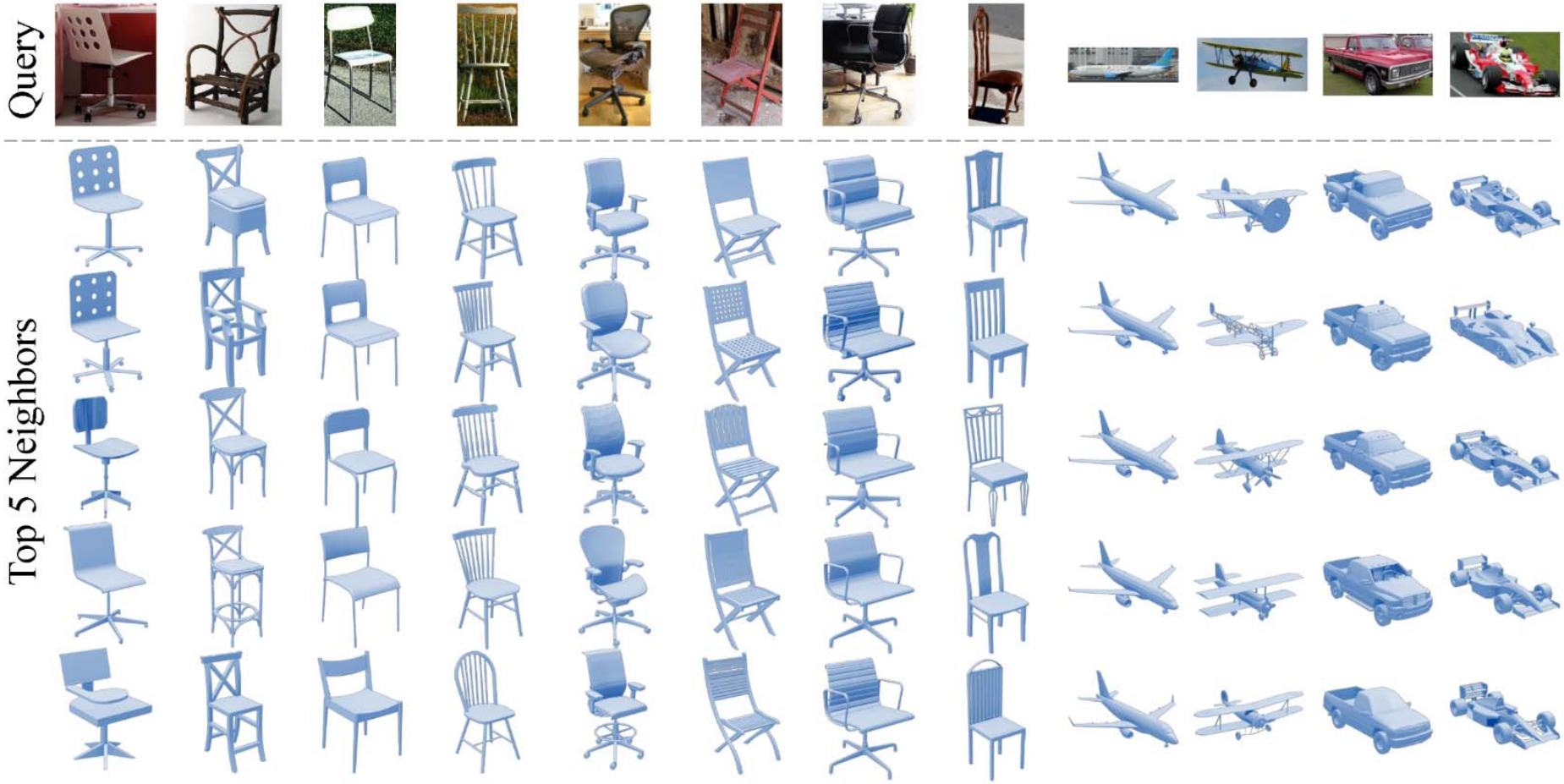
Fig. 2. Schematic of the used architecture.

Combining multiple types of data (2D & 3D)(



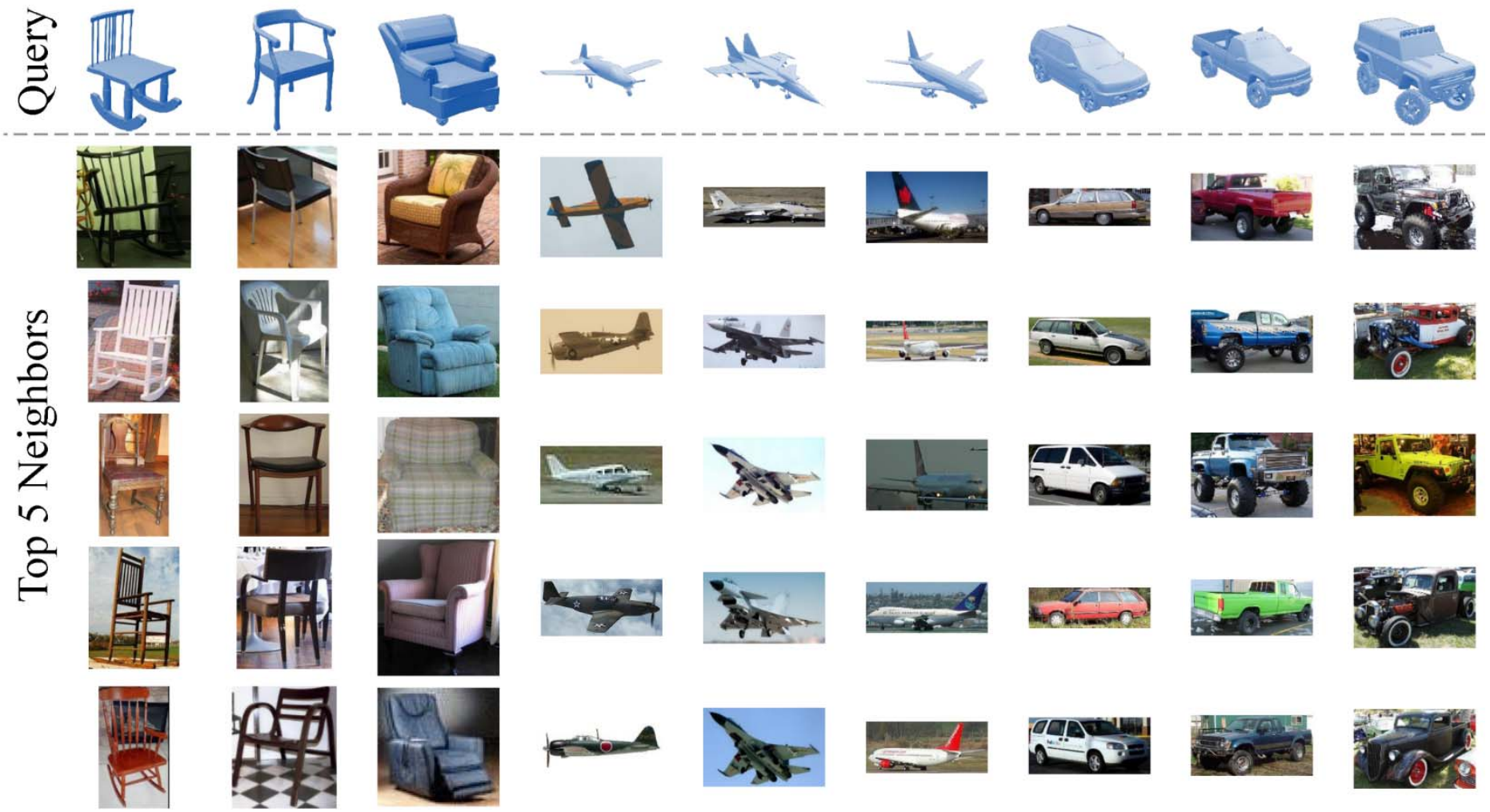
Slides by Li, et al.

Image based Shape Retrieval



Slides by Li, et al.

Shape based Image Retrieval



Slides by Li, et al.

ML and Fairness

- AI effects our lives in many ways
- Widespread algorithms with many small interactions
  - e.g. search, recommendations, social media
- Specialized algorithms with fewer but higher-stakes interactions
  - e.g. medicine, criminal justice, finance
- At this level of impact, algorithms can have unintended consequences
- Low classification error is not enough, need fairness

BUSINESS NEWS OCTOBER 10, 2018 / 3:12 AM / 6 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like

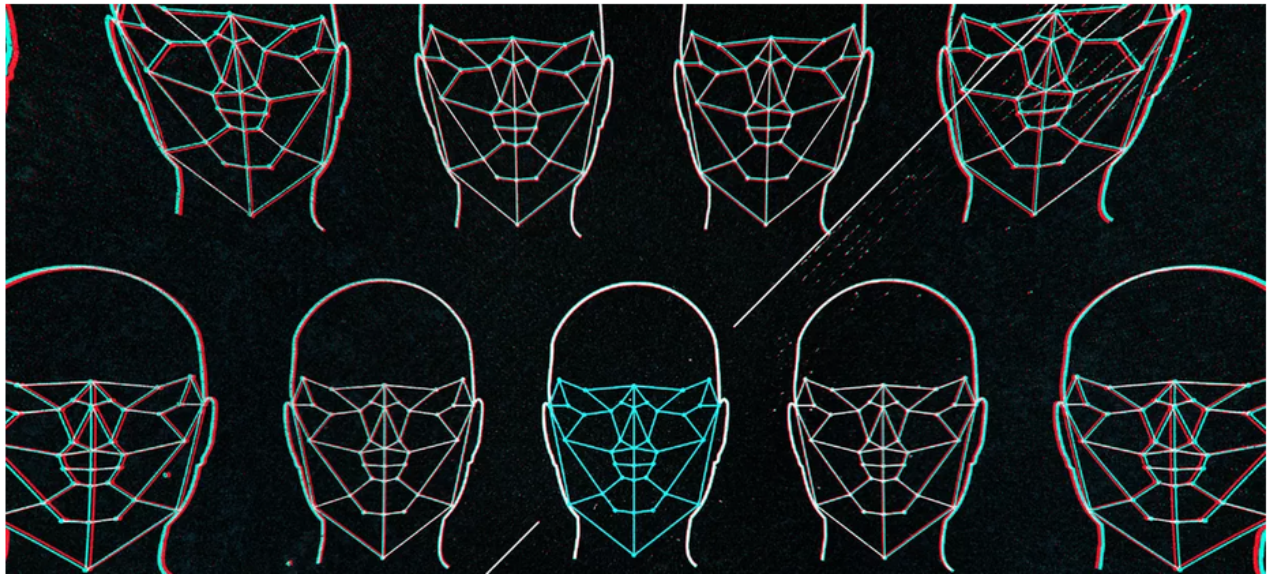
Gender and racial bias found in Amazon's facial recognition technology (again)

17

Research shows that Amazon's tech has a harder time identifying gender in darker-skinned and female faces

By James Vincent | Jan 25, 2019, 9:45am EST

SHARE



MOST READ

My Samsung Galaxy Fold screen broke after just a day

We finally know why the Instagram founders really quit

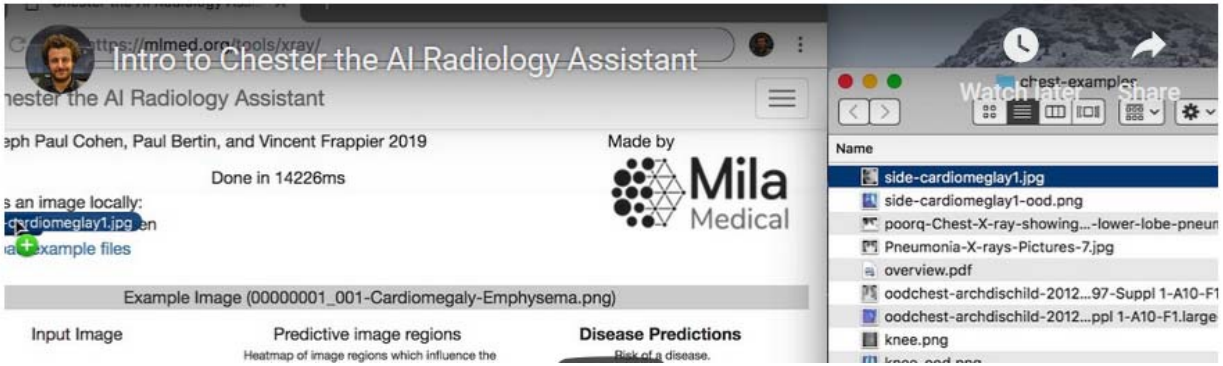
Command Line

Command Line delivers daily updates from the near-future.

Chester the AI Radiology Assistant

NOT FOR MEDICAL USE. This is a web based (but locally run) prototype system for diagnosing chest X-ray images. The patient data remains on your computer and all computation occurs in your browser. The goals of this system are:

1. Let people play with deep learning tools to know how they work and their limitations.
2. Show the potential of open data (needed to build a public system like this).
3. Create a tool to help teach radiology.
4. Demonstrate a model delivery system that can scale to provide free medical tools to the world.



Why is this a problem?

Decision Theory

Define a loss function L(y,a)

y

a

	None	Lung	Breast
Surgery	100	20	10
No surgery	0	50	50

Pick the action with minimum expected loss (risk)

$$a^*(x) = \arg\min_a \sum p(y|x) L(y,a)$$

Equal costs -> Cross Entropy!

ML and Fairness

- Fairness is morally and legally motivated
- Takes many forms
- Criminal justice: recidivism algorithms (COMPAS)
  - Predicting if a defendant should receive bail
  - Unbalanced false positive rates: more likely to wrongly deny a black person bail

Table 1: ProPublica Analysis of COMPAS Algorithm

	White	Black
Wrongly Labeled High-Risk	23.5%	44.9%
Wrongly Labeled Low-Risk	47.7%	28.0%

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Example – Word Embeddings

- Fairness is morally and legally motivated
- Takes many forms
- Bias found in word embeddings (Bolukbasi et al. 2016)
  - Examined word embeddings (word2vec) trained on Google News
  - Represent each word with high-dimensional vector
  - Vector arithmetic: analogies like Paris - France = London - England
  - Found also: man - woman = programmer - homemaker = surgeon - nurse
- The good news: word embeddings learn so well!
- The bad news: sometimes too well
- Our chatbots should be less biased than we are

Slide Credit: David Madras

Example – Word Embeddings

- **Algorithmic fairness**: how can we ensure that our algorithms act in ways that are fair?
  - This definition is vague and somewhat circular
  - Describes a broad set of problems, not a specic technical approach
  - Related to **accountability**: who is responsible for automated behaviour? How do we supervise/audit machines which have large impact?
  - Also **transparency**: why does an algorithm behave in a certain way? Can we understand its decisions? Can it explain itself?
  - Connections to **AI safety** and **aligned AI**: how can we make AI without unintended negative consequences? Aligns with our values?

Why Fairness is Hard

- Suppose we are a bank trying to fairly decide who should get a loan
  - i.e. Who is most likely to pay us back?
- Suppose we have two groups, A and B (the sensitive attribute)
  - This is where discrimination could occur
- The simplest approach is to remove the sensitive attribute from the data, so that our classier doesn't know the sensitive attribute

Table 2: To Loan or Not to Loan?

Age	Gender	Postal Code	Req Amt	A or B?	Pay
46	F	M5E	$300	A	1
24	M	M4C	$1000	B	1
33	M	M3H	$250	A	1
34	F	M9C	$2000	A	0
71	F	M3B	$200	A	0
28	M	M5W	$1500	B	0

Why Fairness is Hard

- However, if the sensitive attribute is correlated with the other attributes, this isn't good enough
- It is easy to predict race if you have lots of other information (e.g. home address, spending patterns)
- More advanced approaches are necessary

Table 3: To Loan or Not to Loan? (masked)

Age	Gender	Postal Code	Req Amt	A or B?	Pay
46	F	M5E	$300	?	1
24	M	M4C	$1000	?	1
33	M	M3H	$250	?	1
34	F	M9C	$2000	?	0
71	F	M3B	$200	?	0
28	M	M5W	$1500	?	0

Definitions of Fairness – Group Fairness

- So we've built our classier . . . how do we know if we're being fair?
- One metric is demographic parity | requiring that the same percentage of A and B receive loans
  - What if 80% of A is likely to repay, but only 60% of B is?
  - Then demographic parity is too strong
- Could require equal false positive/negative rates
  - When we make an error, the direction of that error is equally likely for both groups

$$P(loan|no\ repay, A) = P(loan|no\ repay, B)$$
$$P(no\ loan|would\ repay, A) = P(no\ loan|would\ repay, B)$$

- These are denitions of group fairness
- Treat dierent groups equally"

Definitions of Fairness – Individual Fairness

- Also can talk about individual fairness | \Treat similar examples similarly"
- Learn fair representations
  - Useful for classification, not for (unfair) discrimination
  - Related to domain adaptation
  - Generative modelling/adversarial approaches

(a) Unfair representations (b) Fair(er) representations

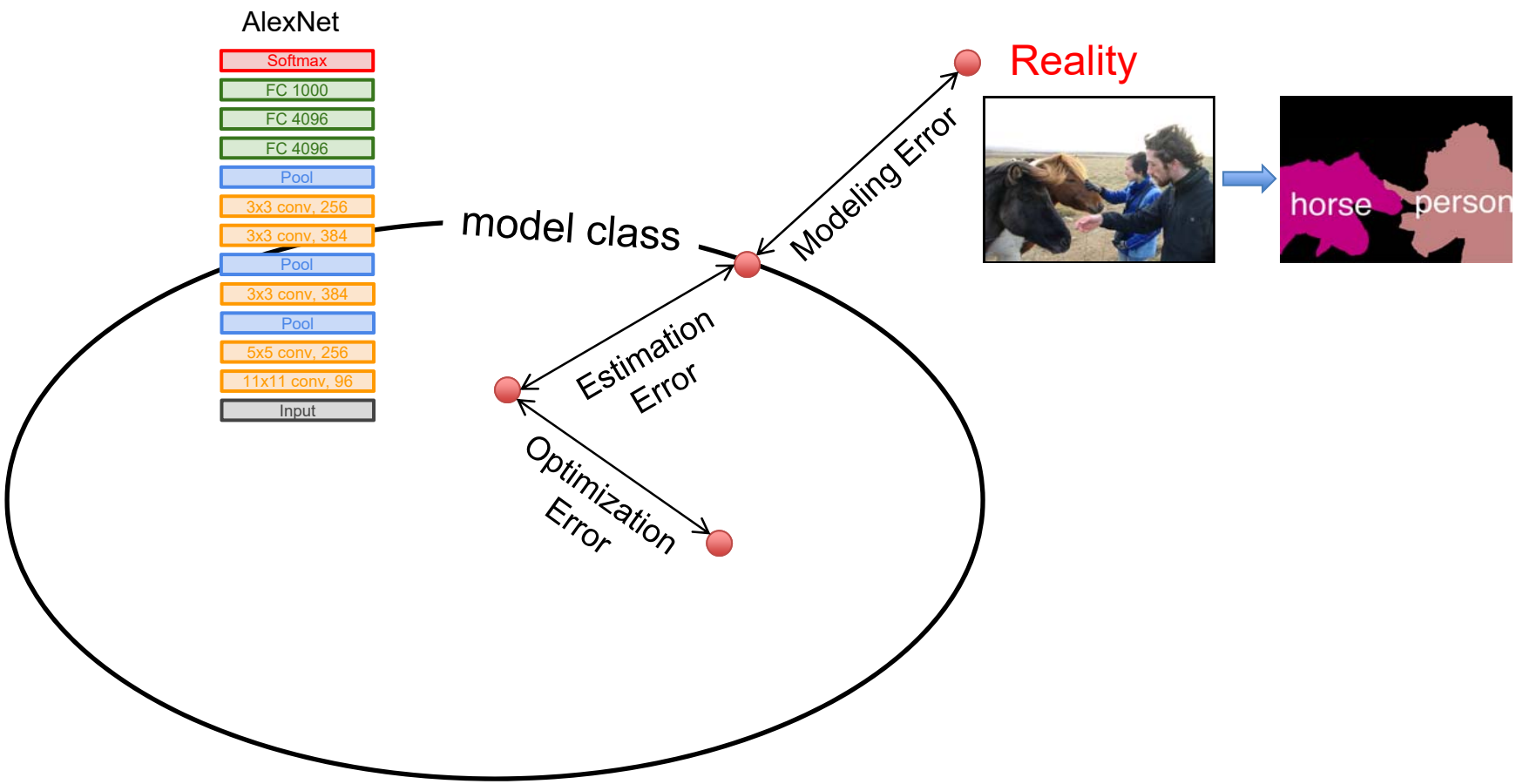
Figure 1: "The Variational Fair Autoencoder" (Louizos et al., 2016)

Conclusion

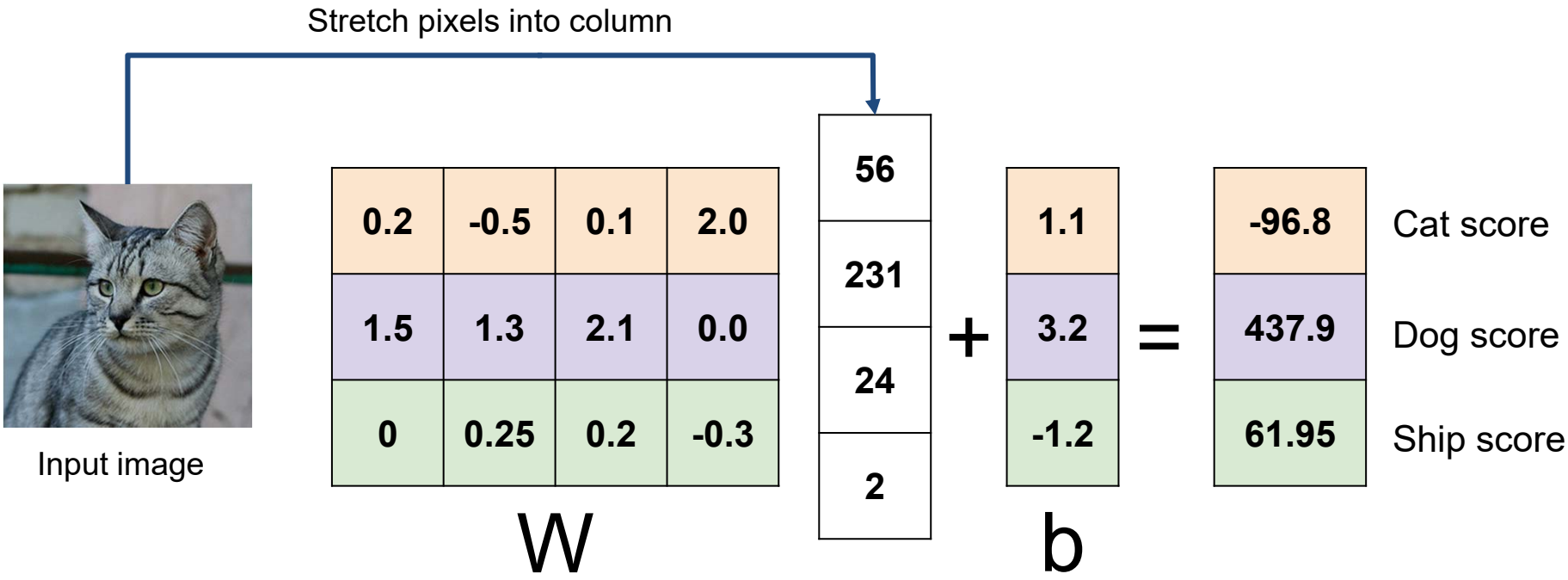
- This is an exciting eld, quickly developing
- Central definitions still up in the air
- AI moves fast | lots of (currently unchecked) power
- Law/policy will one day catch up with technology
- Those who work with AI should be ready

Highlights of Class

Error Decomposition



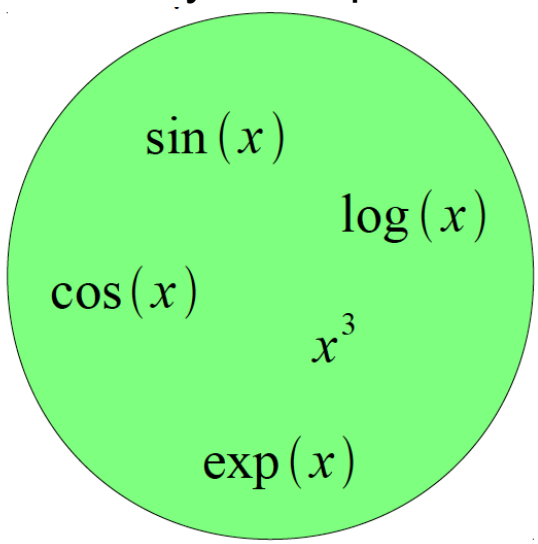
Example with an image with 4 pixels, and 3 classes (cat/dog/ship)



Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

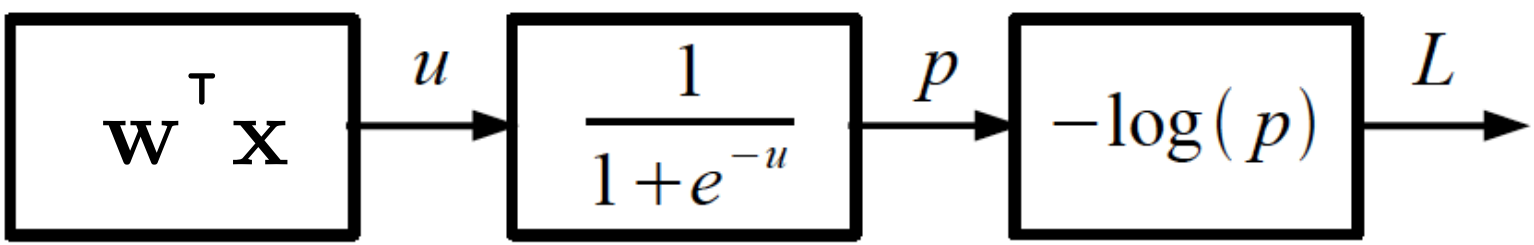
Before: Logistic Regression as Cascade

Given a library of simple functions



Compose into a complicate function

$$-\log\left(\frac{1}{1+e^{-\mathbf{w}^\intercal \mathbf{x}}}\right)$$

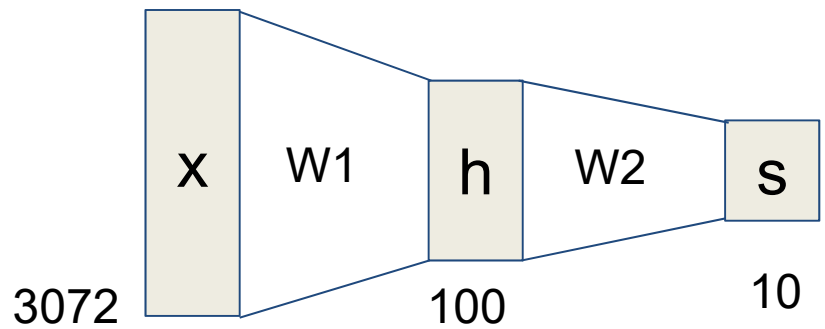


Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Neural networks: without the brain stuff

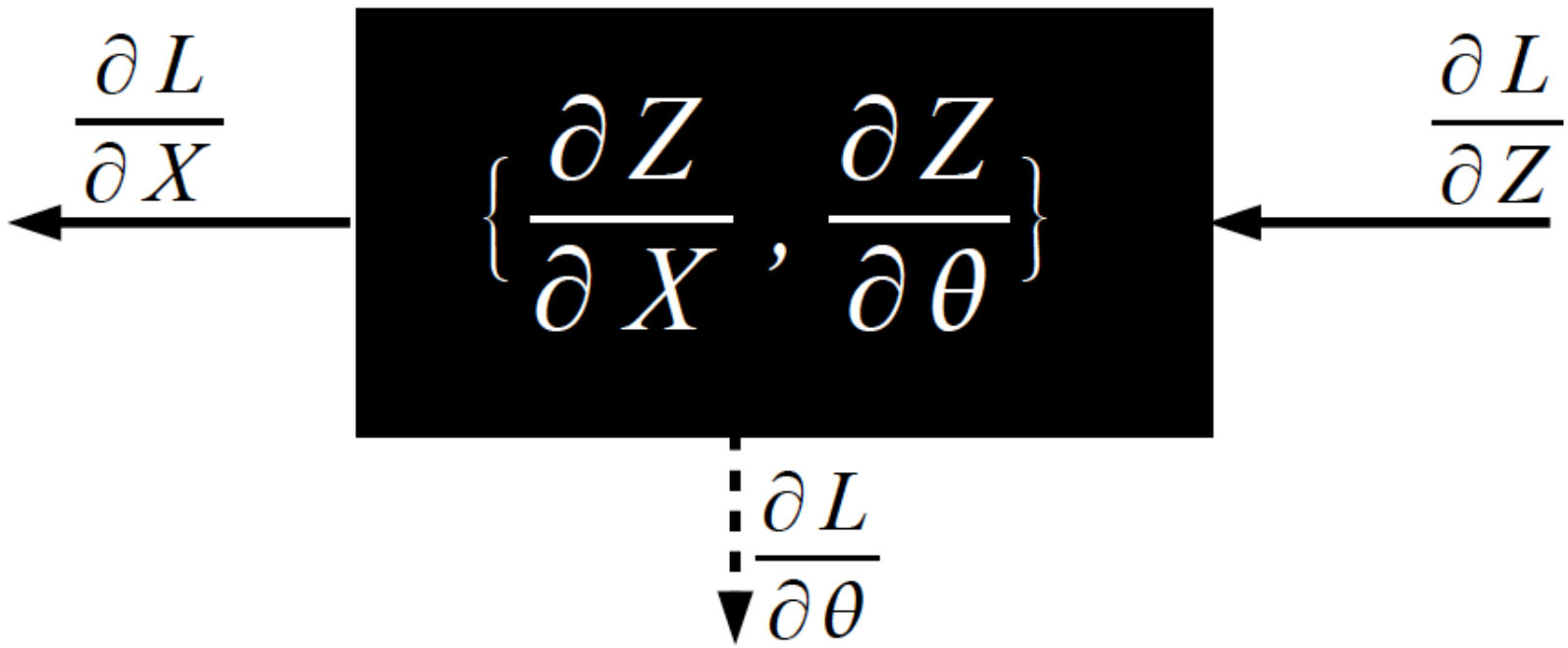
(**Before**) Linear score function: $f = Wx$

(**Now**) 2-layer Neural Network $f = W_2 \max(0, W_1 x)$



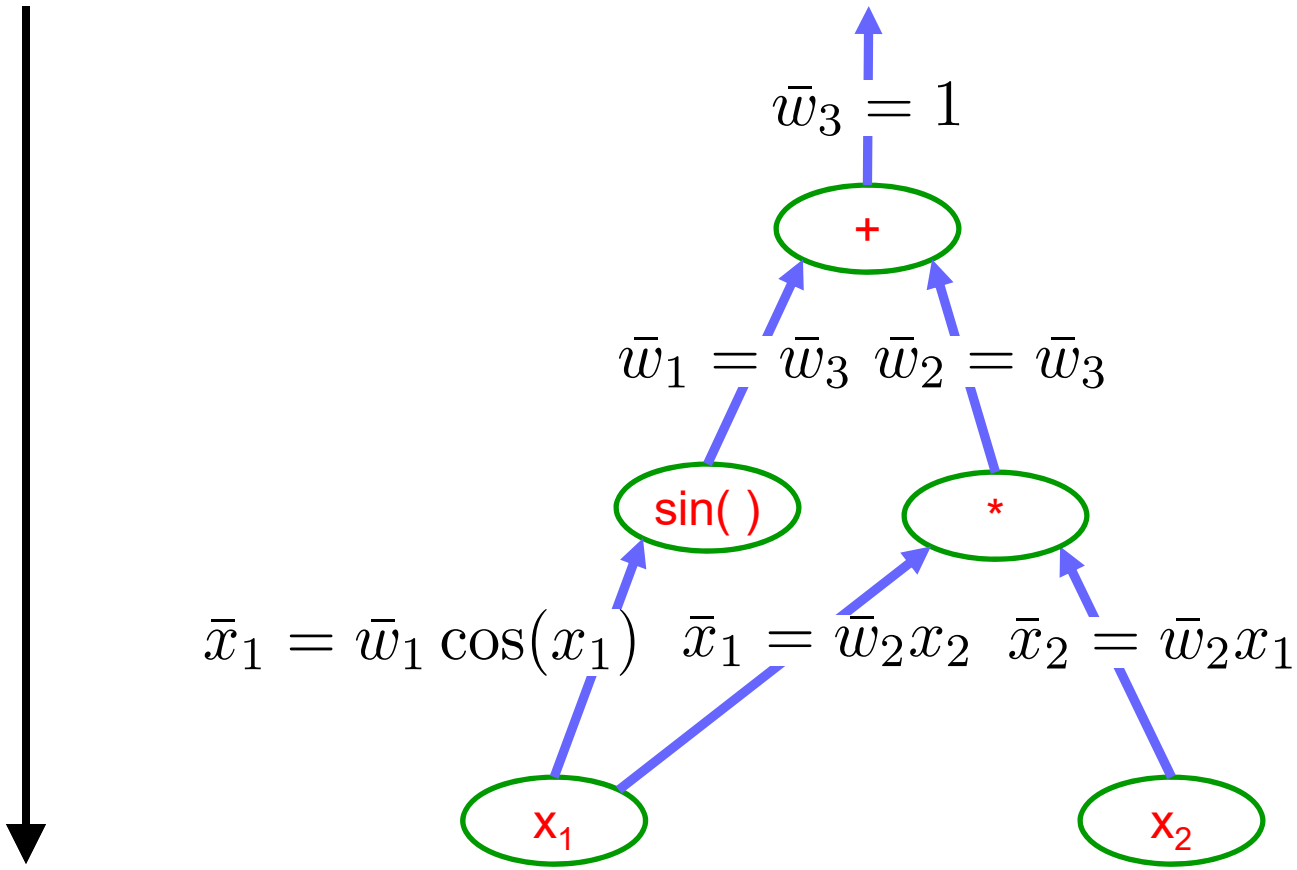
Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

Key Computation: Back-Prop



Example: Reverse mode AD

$$f(x_1, x_2) = x_1 x_2 + \sin(x_1)$$



Adam (full form)

```
first_moment = 0
second_moment = 0
for t in range(1, num_iterations):
  dx = compute_gradient(x)
  first_moment = beta1 * first_moment  + (1 - beta1) * dx
  second_moment = beta2 * second_moment + (1 - beta2) * dx * dx
  first_unbias = first_moment / (1 - beta1 ** t)
  second_unbias = second_moment / (1 - beta2 ** t)
  x -= learning_rate * first_unbias / (np.sqrt(second_unbias) + 1e-7))
```

Momentum

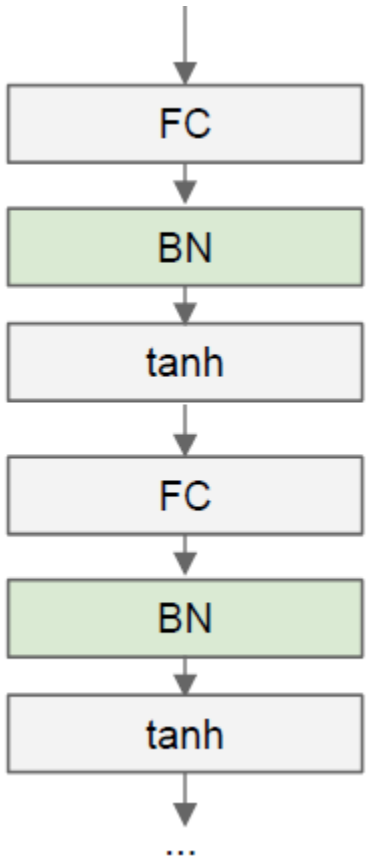
Bias correction

AdaGrad / RMSProp

Bias correction for the fact that first and second moment estimates start at zero

Kingma and Ba, “Adam: A method for stochastic optimization”, ICLR 2015

Batch Normalization



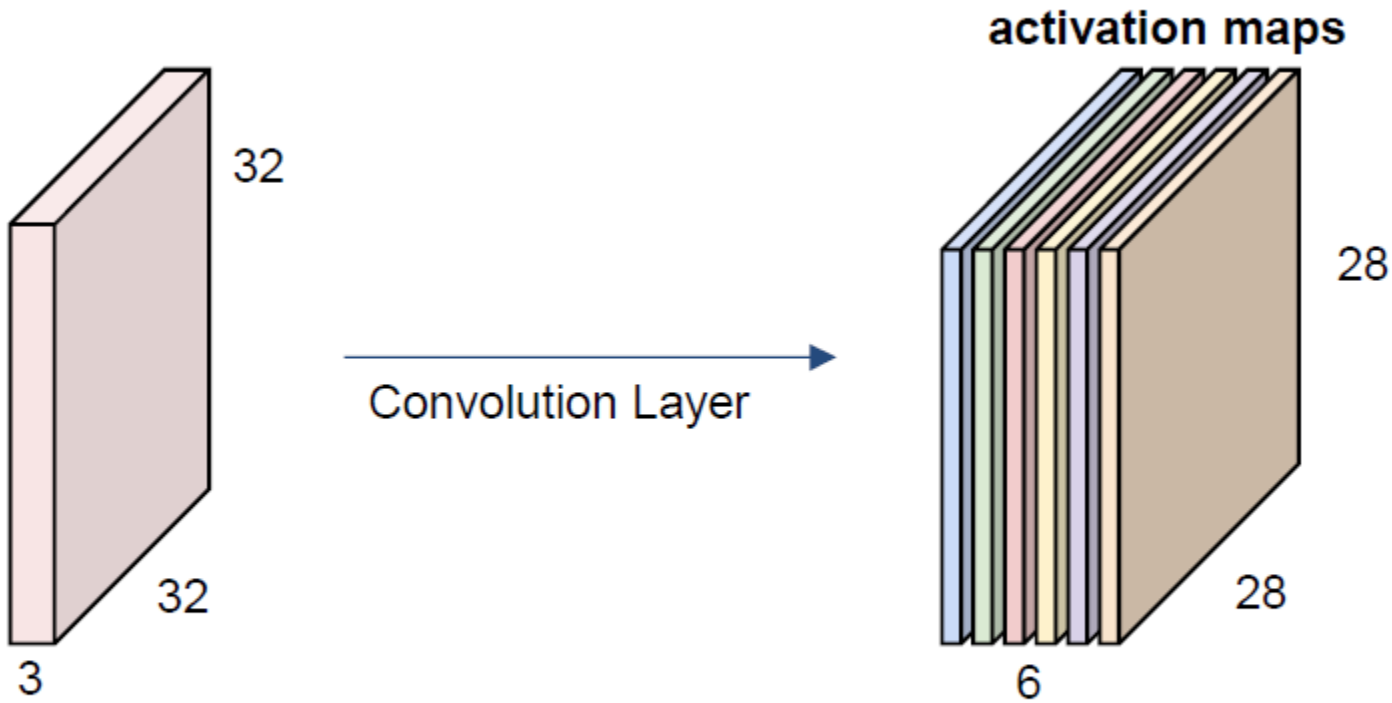
Usually inserted after Fully Connected or Convolutional layers, and before nonlinearity.

Problem: do we necessarily want a zero-mean unit-variance input?

$$\widehat{x}^{(k)} = \frac{x^{(k)} - \mathrm{E}[x^{(k)}]}{\sqrt{\mathrm{Var}[x^{(k)}]}}$$

Convolution Layers

For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:

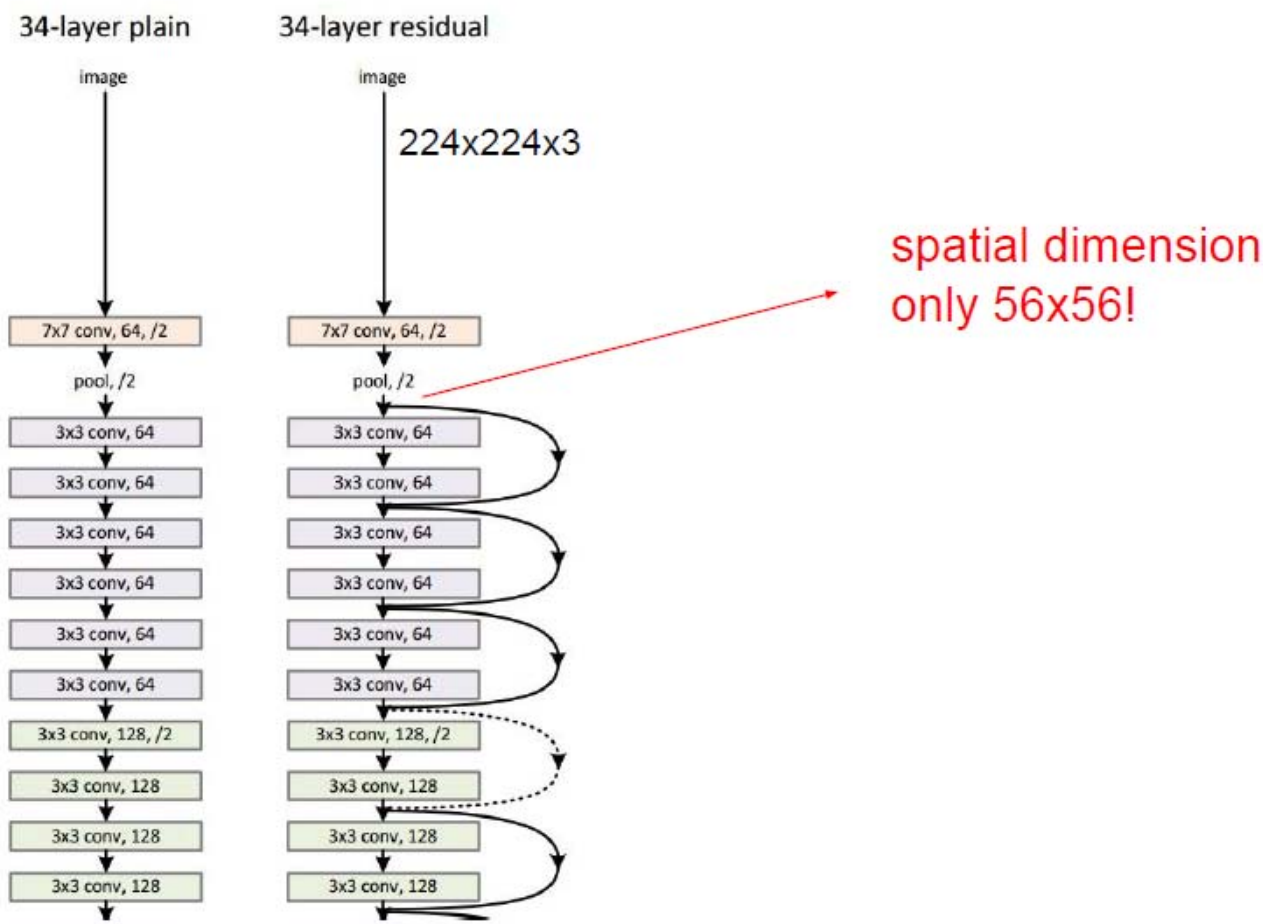


We stack these up to get a "new image" of size 28x28x6!

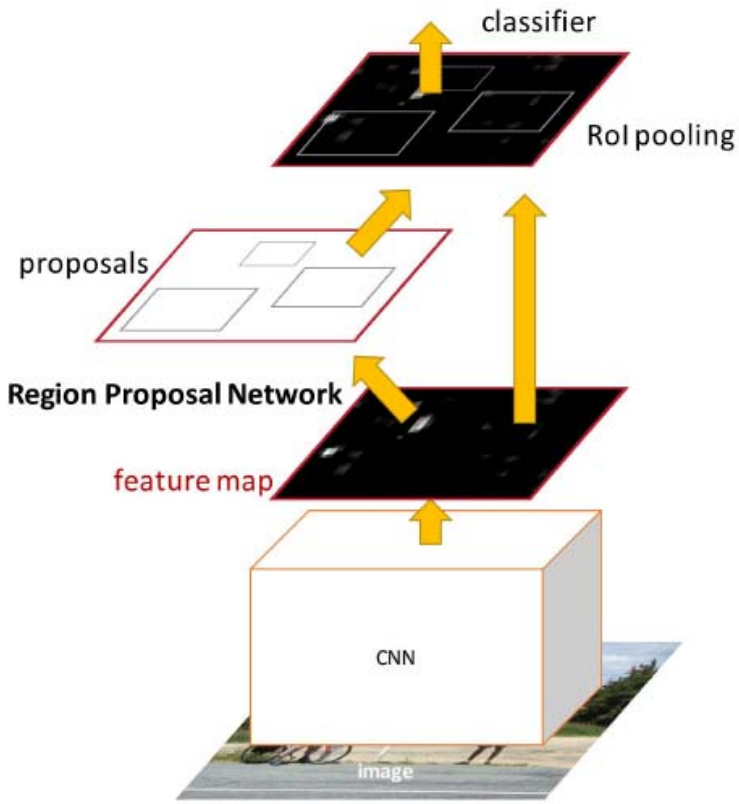
ResNet

Case Study: ResNet

[He et al., 2015]



Faster R-CNN:



Insert a **Region Proposal Network (RPN)** after the last convolutional layer

RPN trained to produce region proposals directly; no need for external region proposals!

After RPN, use RoI Pooling and an upstream classifier and bbox regressor just like Fast R-CNN

Ren et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", NIPS 2015

Slide credit: Ross Girschick

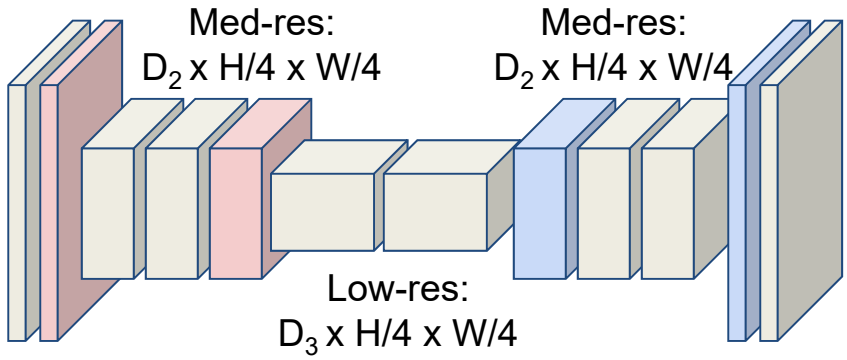
Semantic Segmentation Idea: Fully Convolutional

Downsampling:
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling:
???

Input:
3 x H x W



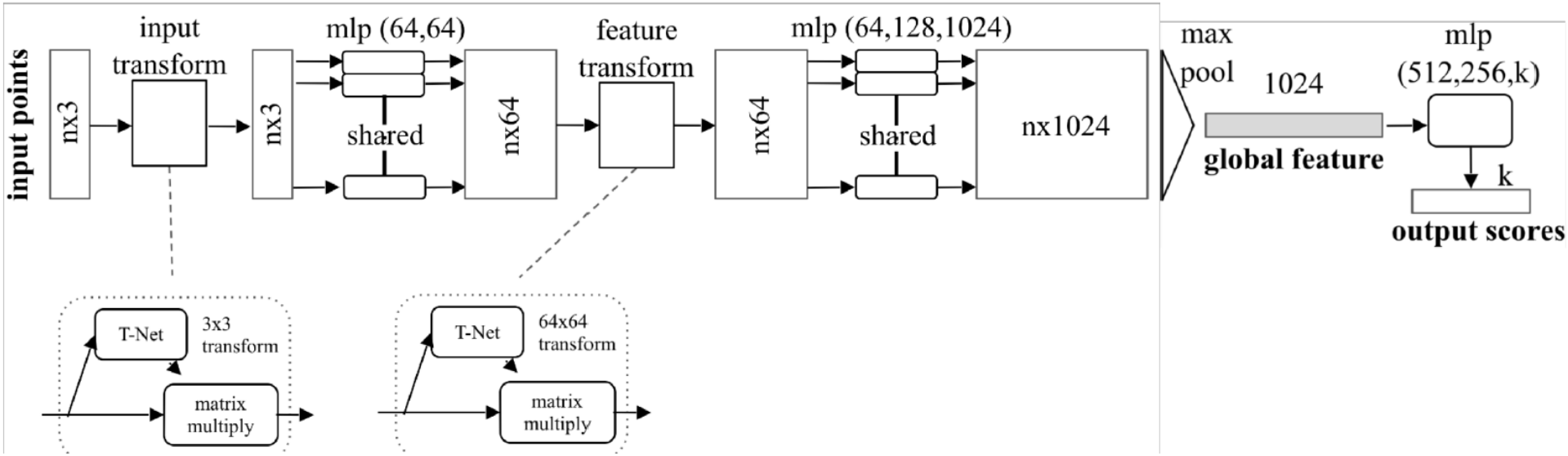
High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

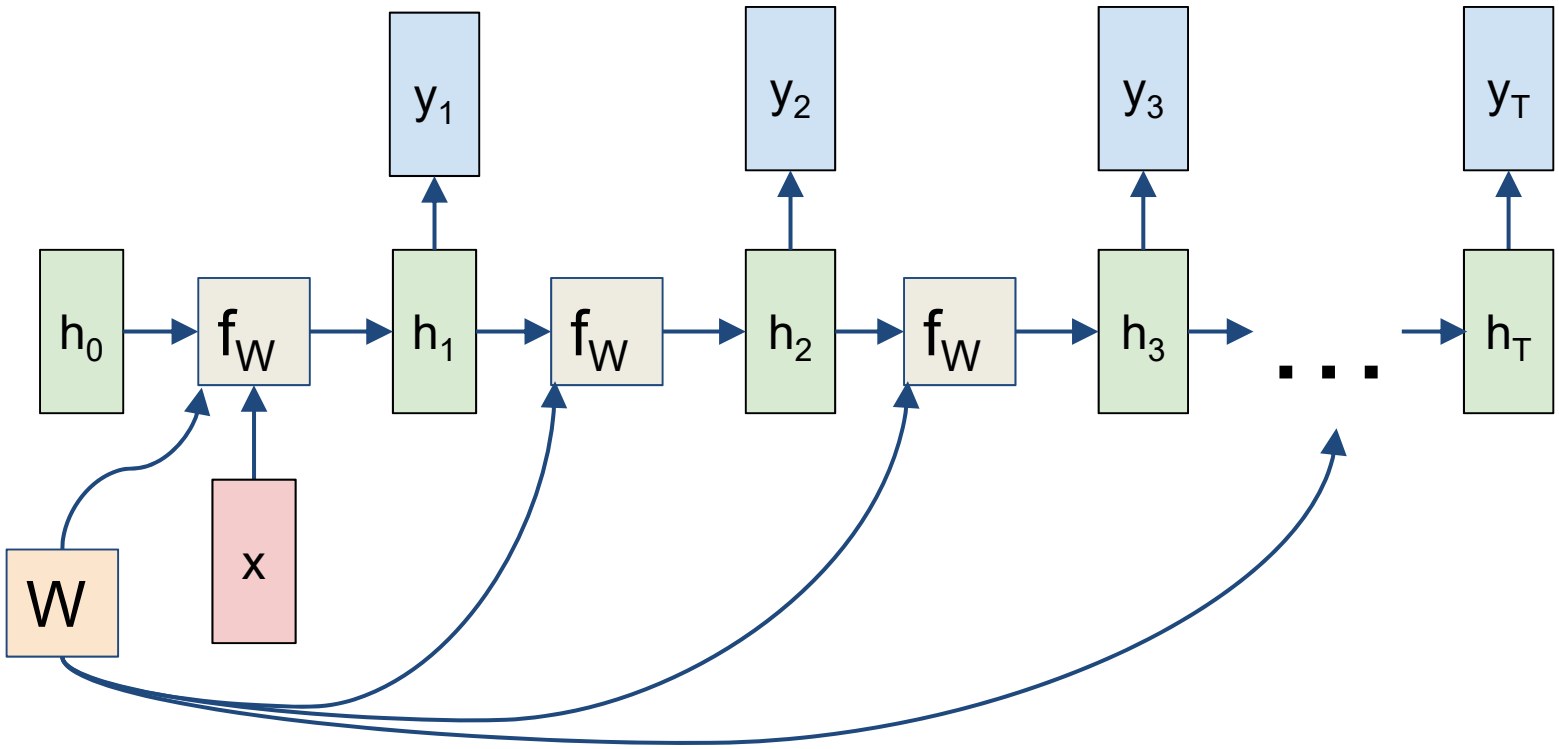
Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

3D CNNs

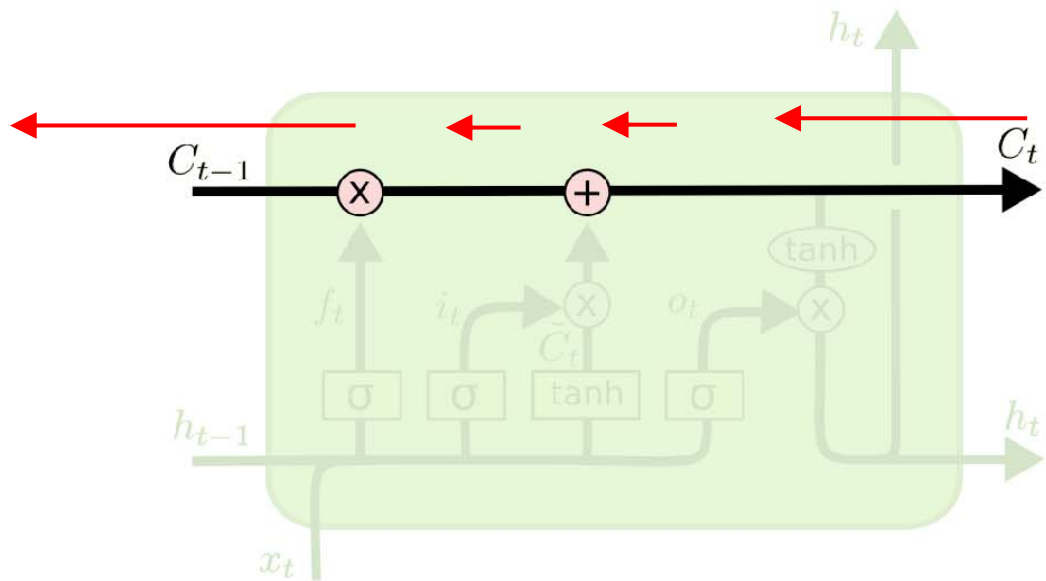

input points
nx3
input transform
nx3
mlp (64,64)
shared
nx64
feature transform
nx64
mlp (64,128,1024)
shared
nx1024
max pool
1024
global feature
mlp (512,256,k)
k
output scores
T-Net
3x3 transform
matrix multiply
T-Net
64x64 transform
matrix multiply

RNN: Computational Graph: One to Many



Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

LSTMs Intuition: Additive Updates



Backpropagation from c_t to c_{t-1} only elementwise multiplication by f, no matrix multiply by W

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Image Credit: Christopher Olah (http://colah.github.io/posts/2015-08-Understanding-LSTMs/)

Unsupervised Learning

Supervised Learning

x	→	Classification	→ y	Discrete
x	→	Regression	→ y	Continuous

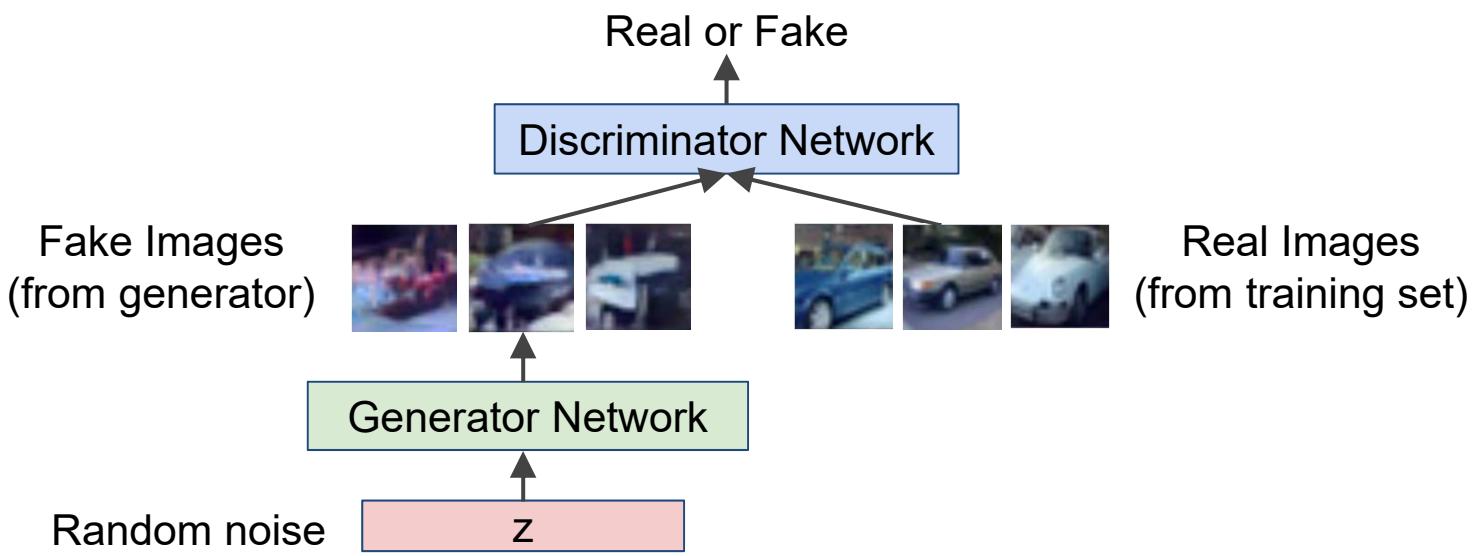
Unsupervised Learning

x	→	Clustering	→ c	Discrete
x	→	Dimensionality Reduction	→ z	Continuous
x	→	Density Estimation	→ p(x)	On simplex

Training GANs: Two-player game

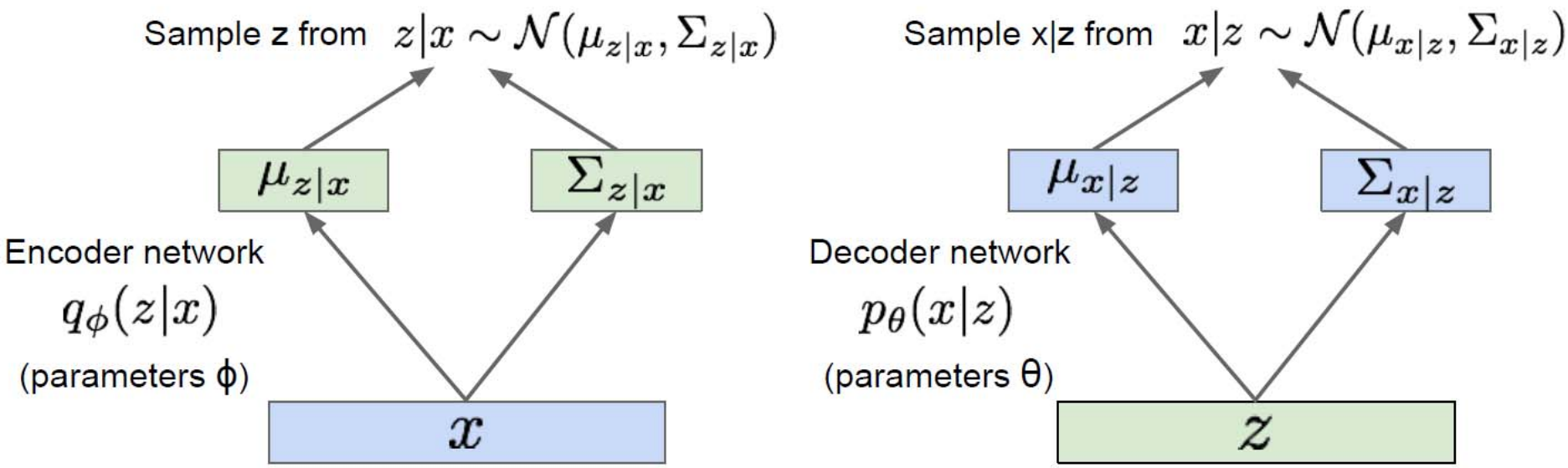
Ian Goodfellow et al., “Generative Adversarial Nets”, NIPS 2014

Generator network: try to fool the discriminator by generating real-looking images
Discriminator network: try to distinguish between real and fake images



Variational Autoencoders: Intractability

Since we're modeling probabilistic generation of data, encoder and decoder networks are probabilistic



Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

[Mnih et al. NIPS Workshop 2013; Nature 2015]

Q-network Architecture

$Q(s, a; \theta)$:
neural network
with weights θ

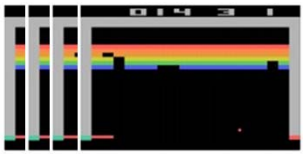
A single feedforward pass to compute Q-values for all actions from the current state => efficient!

FC-4 (Q-values)

FC-256

32 4x4 conv, stride 2

16 8x8 conv, stride 4



Last FC layer has 4-d output (if 4 actions), corresponding to $Q(s_t, a_1)$, $Q(s_t, a_2)$, $Q(s_t, a_3)$, $Q(s_t, a_4)$

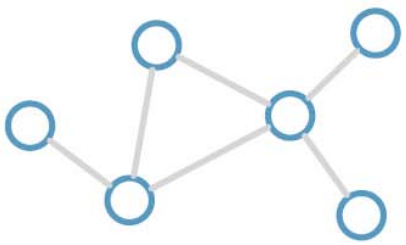
Number of actions between 4-18 depending on Atari game

Current state s_t: 84x84x4 stack of last 4 frames
(after RGB->grayscale conversion, downsampling, and cropping)

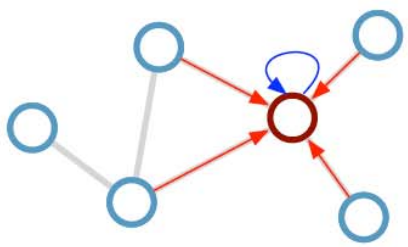
Graph convolutional networks (GCNs)

Kipf & Welling (ICLR 2017), related previous works by Duvenaud et al. (NIPS 2015) and Li et al. (ICLR 2016)

Consider this undirected graph:



Calculate update for node in red:



Update rule:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{h}_i^{(l)}\mathbf{W}_0^{(l)} + \sum_{j\in\mathcal{N}_i} \frac{1}{c_{ij}}\mathbf{h}_j^{(l)}\mathbf{W}_1^{(l)}\right)$$

Scalability: subsample messages [Hamilton et al., NIPS 2017]

Desirable properties:

- Weight sharing over all locations
- Invariance to permutations
- Linear complexity O(E)
- Applicable both in transductive and inductive settings

Limitations:

- Requires gating mechanism / residual connections for depth
- Only indirect support for edge features

$\mathcal{N}_i$: neighbor indices

c_{ij} : norm. constant (fixed/trainable)

We have come a long way

- **ML background** – error decomposition, overfitting, features, etc.
- **Linear classifiers**
  - **& softmax**
- **Computation Graph**
- **Gradient Descent**
- **Adding layers**
- **Backpropagation, automatic differentiation**
- **Optimization –** regularization/normalization (batch norm, dropout), augmentation, different optimizers (adam, adagrad, etc.)
- **Convolution and Pooling layers**
- **Modern CNNs** - AlexNet, VGG, Inception, ResNet
- **3D CNNs**
- **Recurrent Neural Networks and LSTMs**
  - NLP, word/sentence vectors, attention, etc.
- **Unsupervised feature learning**
- **Generative models (GANs, VAEs)**
- **Deep Reinforcement Learning**
- **Other applications:** Few-Shot Learning, structure

Things to Watch out For

- Research is cyclical
  - SVMs, boosting, probabilistic graphical models & Bayes Nets, Structural Learning, Sparse Coding, Deep Learning
  - Deep learning is unique in its depth and breadth, but...
  - Deep learning may be improved, reinvented, combined, overtaken
- Learn fundamentals for techniques across the field:
  - Know the span of ML techniques and choose the ones that fit your problem!
  - **Be responsible** in 1) how you use it, 2) promises you make and how you convey it
- Try to understand landscape of the field
  - Look out for what is coming up next, not where we are
- Have fun!

Some current/upcoming topics

- Current / Recent Past
  - AutoML
  - Meta-learning
  - Unsupervised, semi-supervised, domain adaptation, zero/one/few-shot learning
  - Memory
  - Visual question answering, embodied question answering
  - Adversarial Examples

- More recent
  - Continual/lifelong learning without forgetting
  - World modeling, learning intuitive/physics models
  - Visual dialogue, agents, chatbots
  - Fixing reinforcement learning
    - First you have to admit you have a problem
  - Simulation frameworks, joint perception, planning, and action
    - Navigation, mapping
  - Deep Learning and logic!
  - Just scaling everything up and watch the magic!

CIOS!

http://gatech.smartevals.com/