Learning Methods

Dual, Self-supervised, Self-augmented Learnings

Hao Dong

2019, Peking University

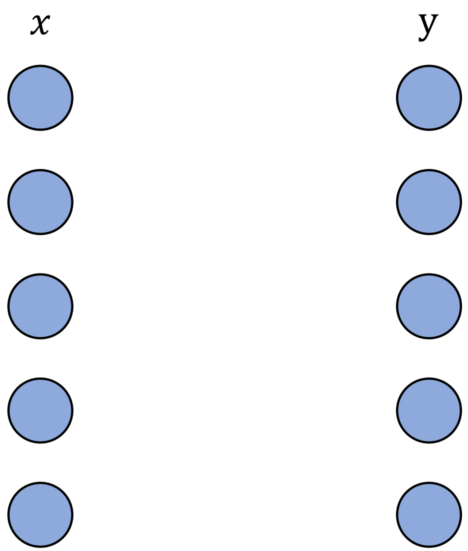
Learning Methods

- Dual, Self-supervised, Self-augmented Learnings
- Dual Learning
- Self-supervised Learning
- Self-augmented Learning
- Summary

From Mapping Point of View
Dual, Self-supervised, Self-augmented Learning

From Mapping Point of View

Data in both input and output
(Learn the mapping f, f')

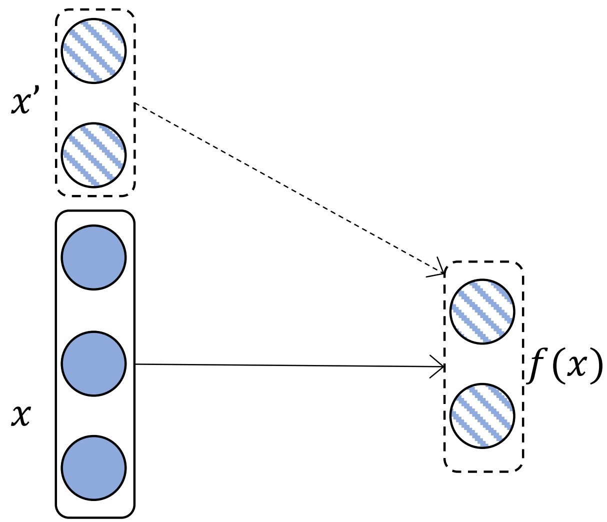


$y = f(x), x = f'(y)$

(Unsupervised) Dual Learning

- VAE
- CycleGAN
- …

Data in input x, x' only
with known mapping f'
(Learn the mapping f)

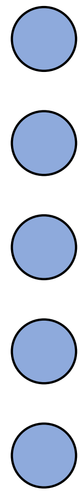


$x' = f(x)$

Self-supervised Learning

- Word2Vec
- Denoising Autoencoder
- …

Data in input only
with known inverse mapping f'
(Learn the mapping f and output y)



$y = f(x), x = f'(y)$

Self-augmented Learning

- ?

Dual Learning

Dual Learning

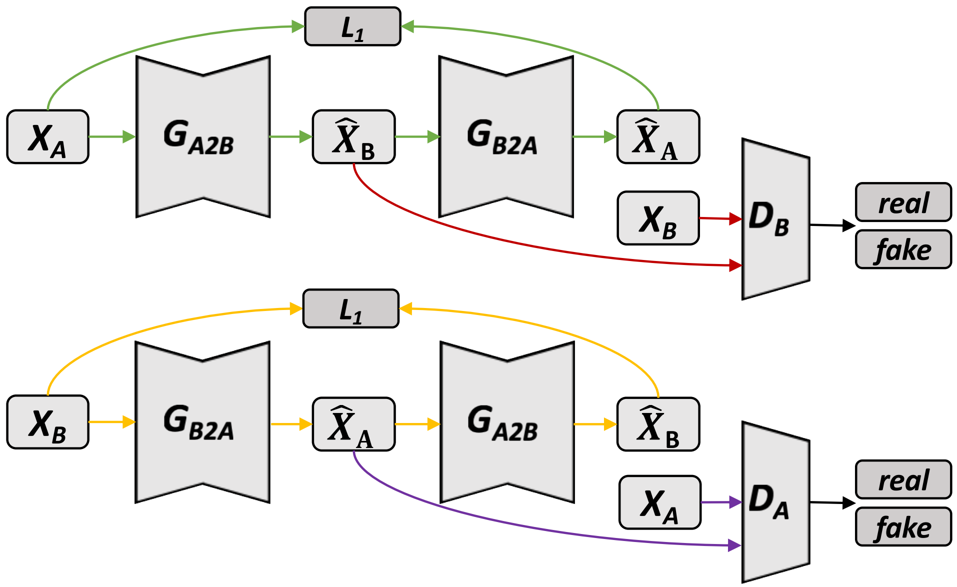
北京大學 PEKING UNIVERSITY

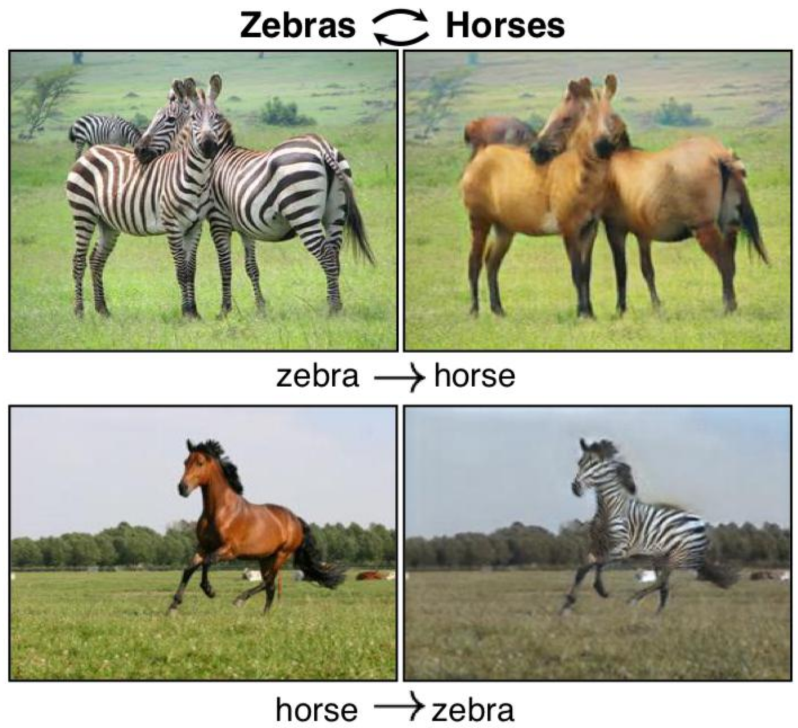
- Motivation
  - Human label is expensive
  - No feedback if using unlabeled data

Application	Primal Task	Dual (Inverse) Task
Machine translation	Translate language from A to B	Translate language from B to A
Speed processing	Speech to text (STT)	Text to speech (TTS)
Image understanding	Image captioning	Image generation
Conversation engine	Question	Answer
Search engine	Search	Query

Dual Learning

- Example: Unpaired Image-to-Image Translation

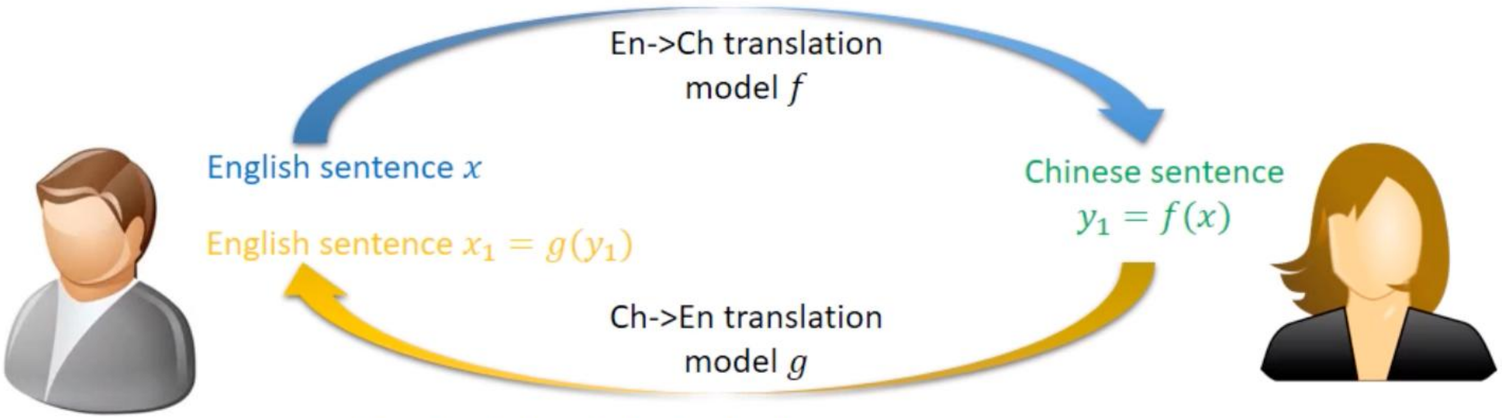




Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

Dual Learning

- Example: Language Translation



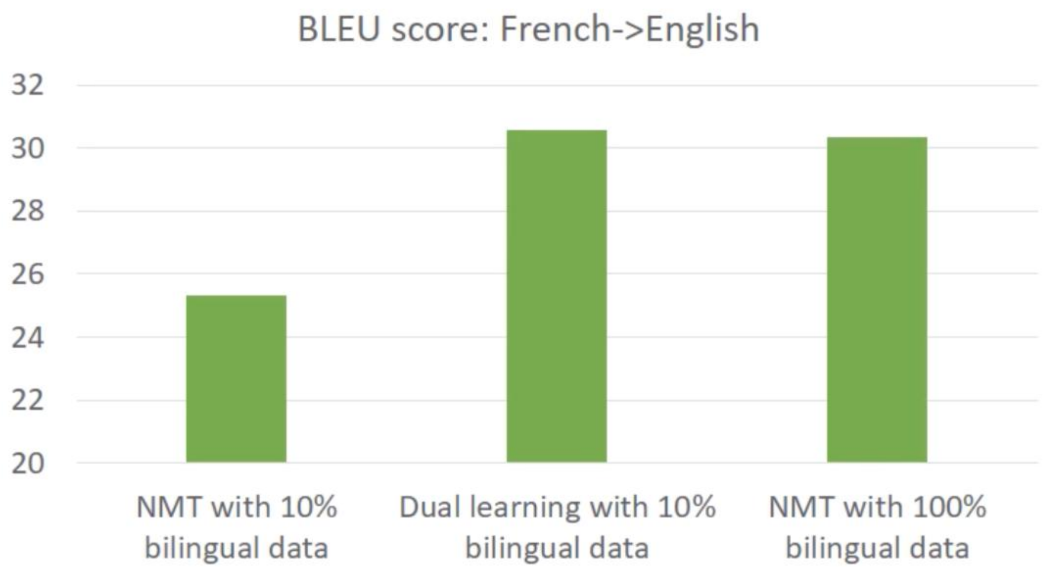
Feedback signals during the loop:
- $s(x, x_1)$: BLEU score of x_1 given x
- $L(y)$ and $L(x_1)$: Likelihood and language model of y_1 and x_1

Reinforcement learning is used to improve the translation models from these feedback signals

Dual Learning for Machine Translation.
Yingce Xia, Di He, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, Wei-Ying Ma. NIPS, 2016

Dual Learning

- Example: Machine Translation



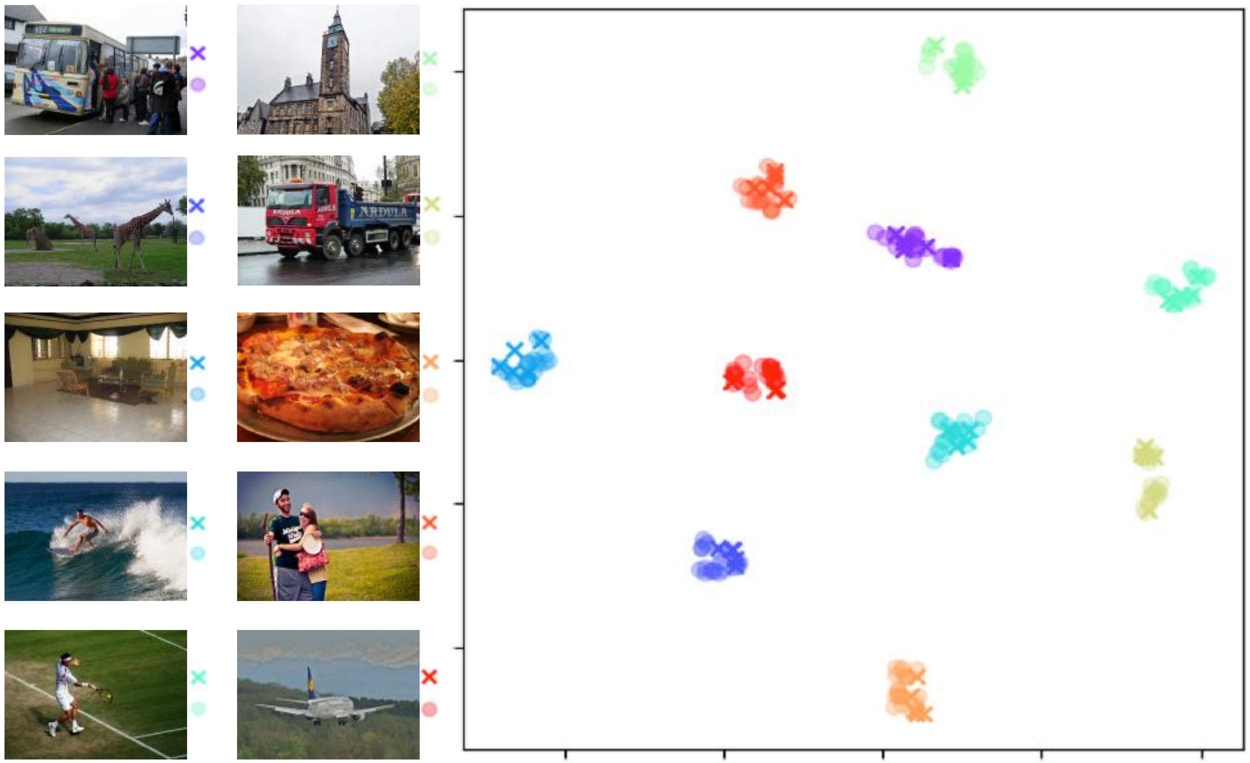
Starting from initial models obtained from only 10% bilingual data, dual learning can achieve similar accuracy as the NMT model learned from 100% bilingual data!

Dual Learning for Machine Translation.
Yingce Xia, Di He, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, Wei-Ying Ma. NIPS, 2016

Dual Learning

北京大學 PEKING UNIVERSITY

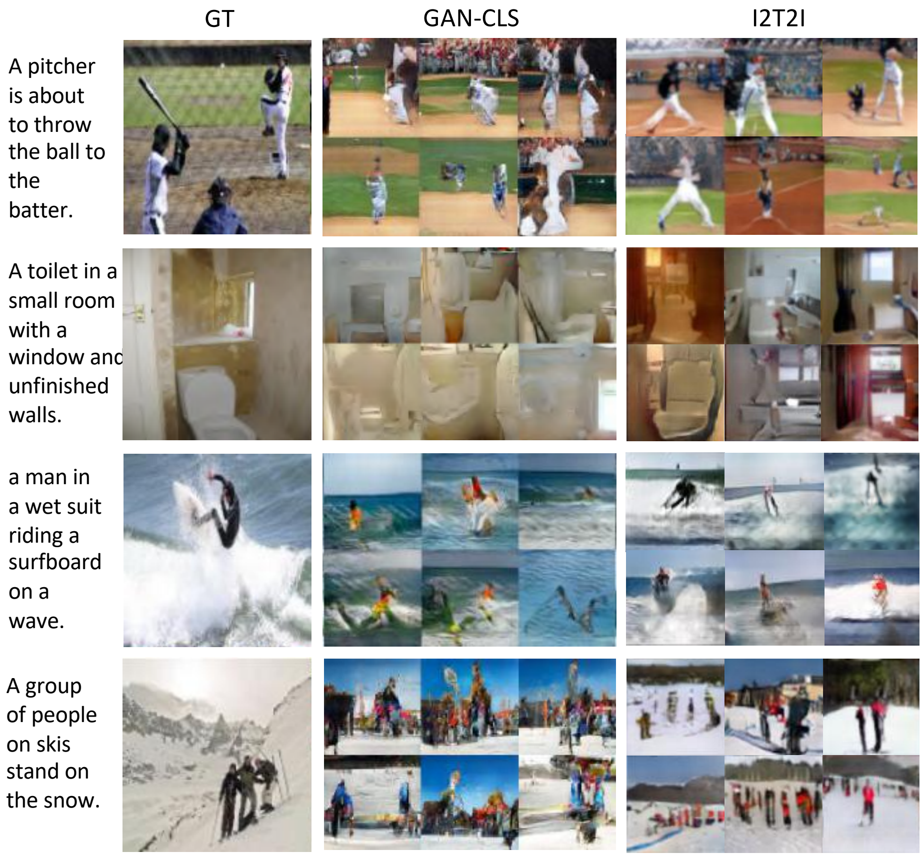
- Example: Image-to-Text-to-Image, I2T2I



I2T2I: Learning Text to Image Synthesis with Textual Data Augmentation. *Hao Dong, Simiao Yu, etal*. ICIP, 2017.

Dual Learning

- Example: Image-to-Text-to-Image, I2T2I



I2T2I: Learning Text to Image Synthesis with Textual Data Augmentation. *Hao Dong, Simiao Yu, etal*. ICIP, 2017.

Self-supervised Learning

Self-supervised Learning

- Self-supervised learning is autonomous supervised learning, it learns to predict part of its input from other parts of its input.

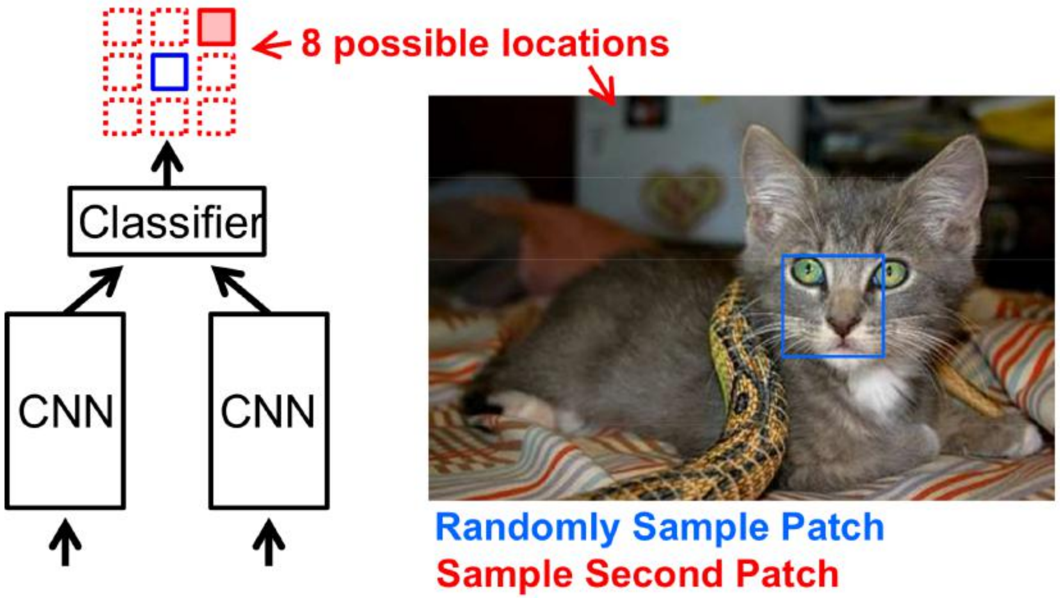
- Examples: Word2Vec, Denoising Autoencoder

- Self-supervised vs. unsupervised learning: Self-supervised learning is like unsupervised Learning because the system learns without using explicitly-provided labels. It is different from unsupervised learning because we are not learning the inherent structure of data. Self-supervised learning, unlike unsupervised learning, is not centered around clustering and grouping, dimensionality reduction, recommendation engines, density estimation, or anomaly detection.

Self-supervised Learning

- **Image Example**: Relative Positioning

Train network to predict relative position of two regions in the same image

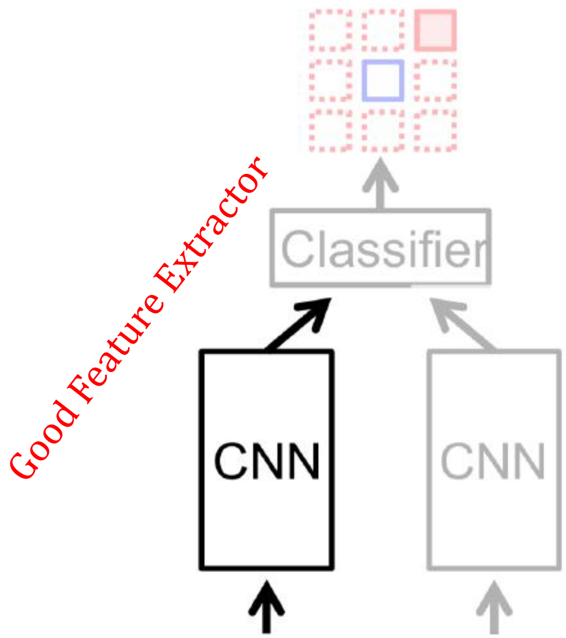


Unsupervised visual representation learning by context prediction, Carl Doersch, Abhinav Gupta, Alexei A. Efros, ICCV 2015

Self-supervised Learning

- Image Example: Relative Positioning

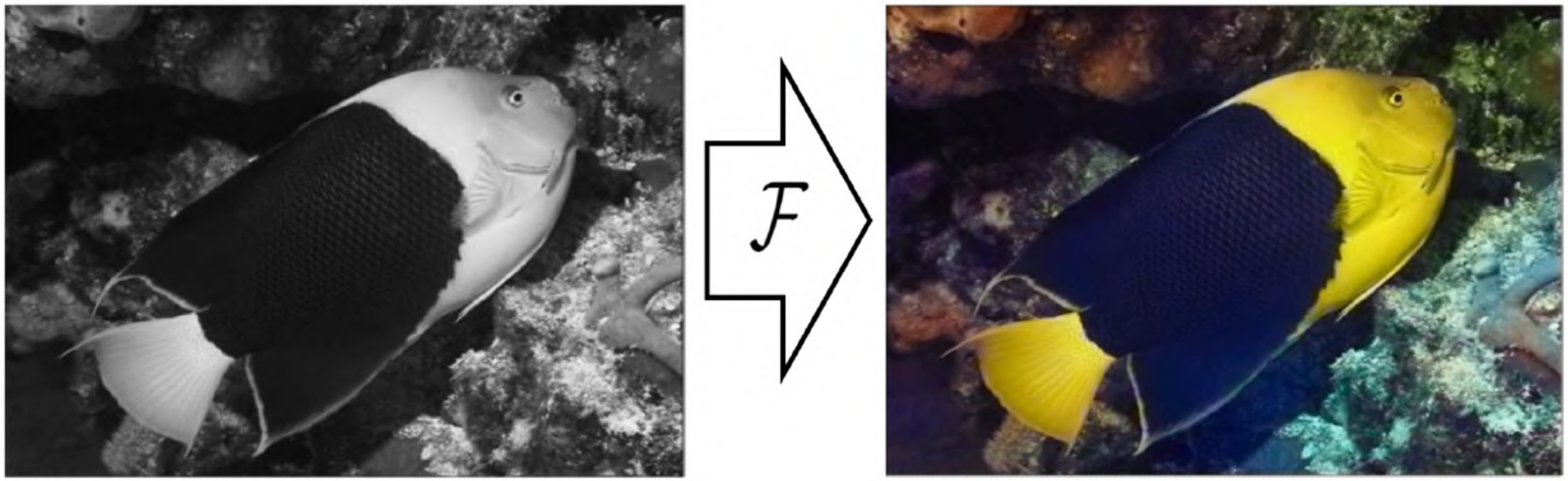
Learn high-level features



Unsupervised visual representation learning by context prediction, Carl Doersch, Abhinav Gupta, Alexei A. Efros, ICCV 2015

Self-supervised Learning

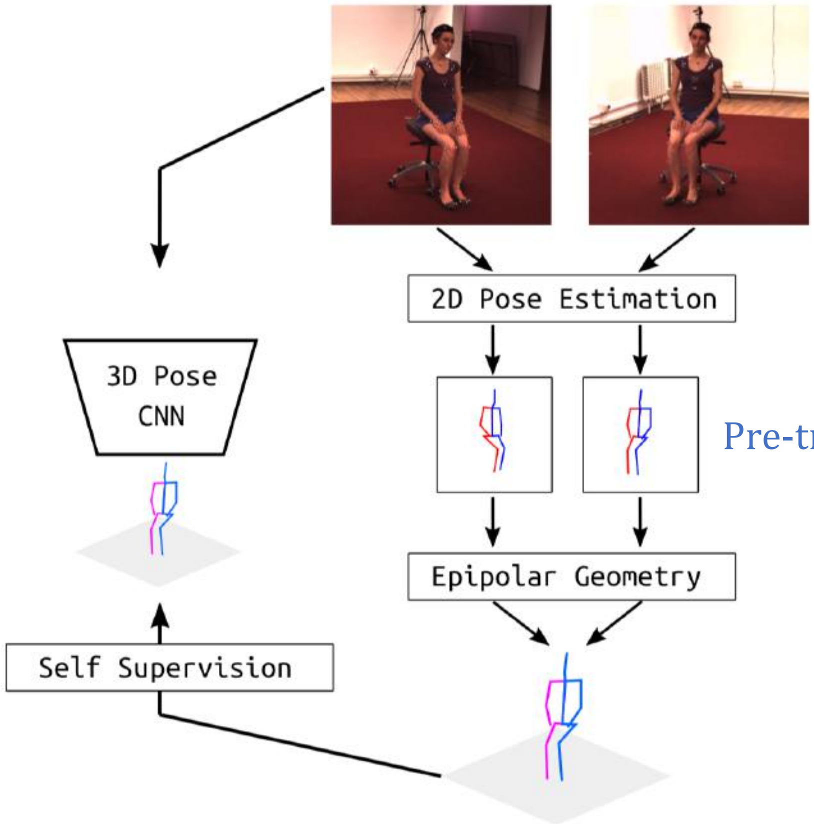
- Image Example: Colourization



Colorful Image Colorization. *Zhang et al.,* ECCV 2016

Self-supervised Learning

- Image Example: 3D pose estimation



Pre-trained on MPII

Self-Supervised Learning of 3D Human Pose using Multi-view Geometry. *Muhammed Kocabas, Salih Karagoz, Emre Akbas.* CVPR, 2019

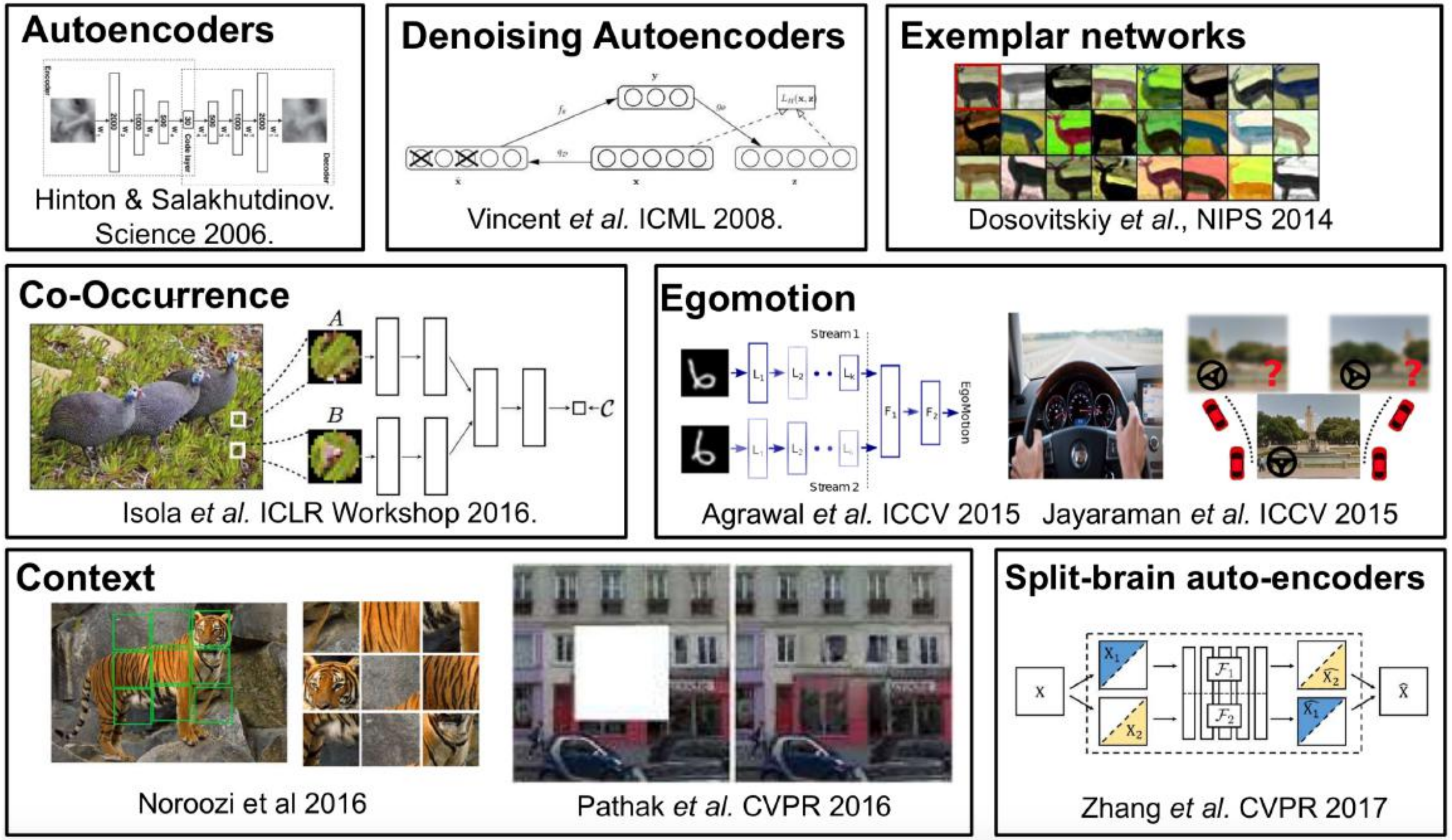
Self-supervised Learning

- Image Example: Learn from Rotation

Unsupervised representation learning by predicting image rotations, *Spyros Gidaris, Praveer Singh, Nikos Komodakis,* ICLR 2018

Self-supervised Learning

- Image Examples



Self-supervised Learning

- **Video Example**

- Videos contain
  - Color, Temporal info
- Possible proxy tasks
  - Temporal order of the frames
  - Optical flow: Motion of objects
  - …

Self-supervised Learning

- Video Example: Shuffle and Learn

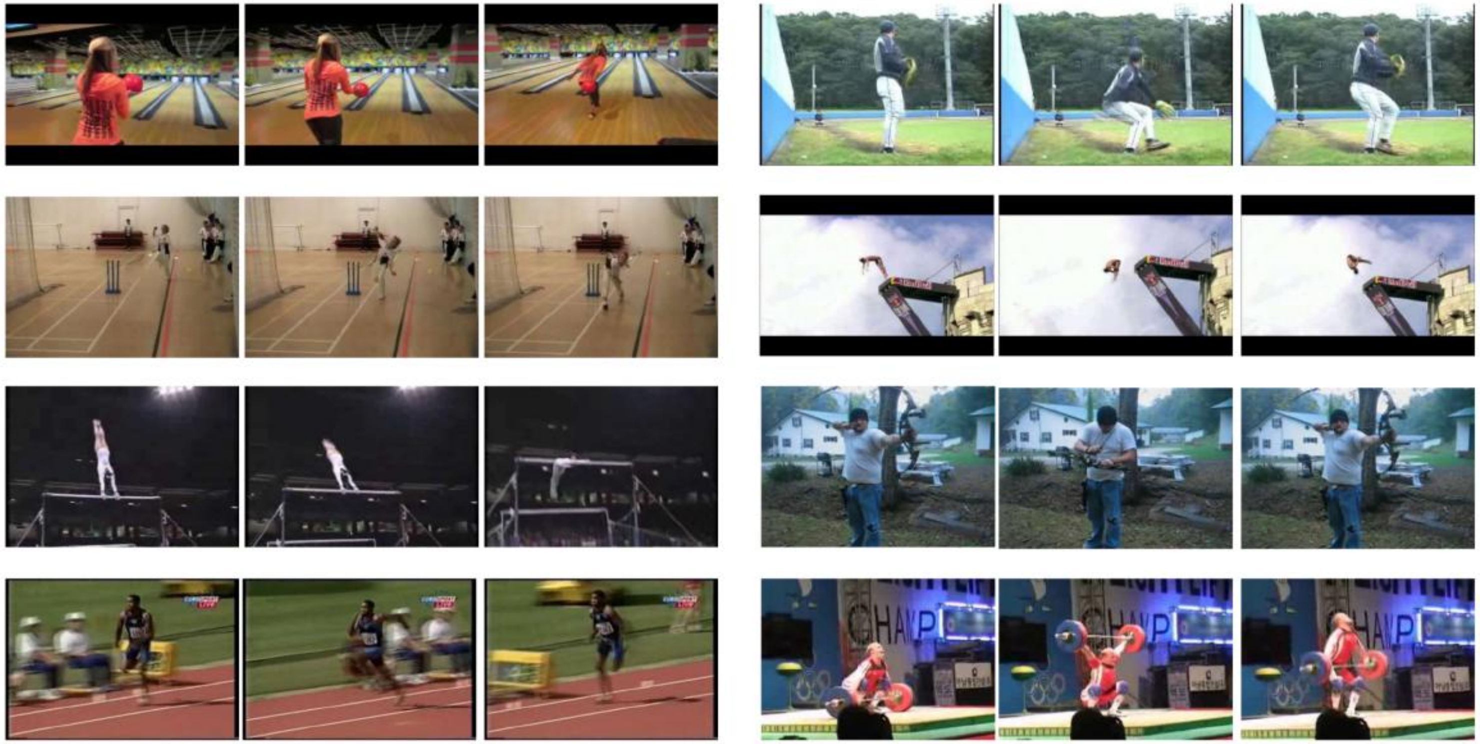
Given a start and an end, can this point lie in between?

Unsupervised Learning using Temporal Order Verification. *Ishan Misra, C. Lawrence Zitnick and Martial Hebert.* ECCV, 2016.

Self-supervised Learning

- Video Example: Shuffle and Learn

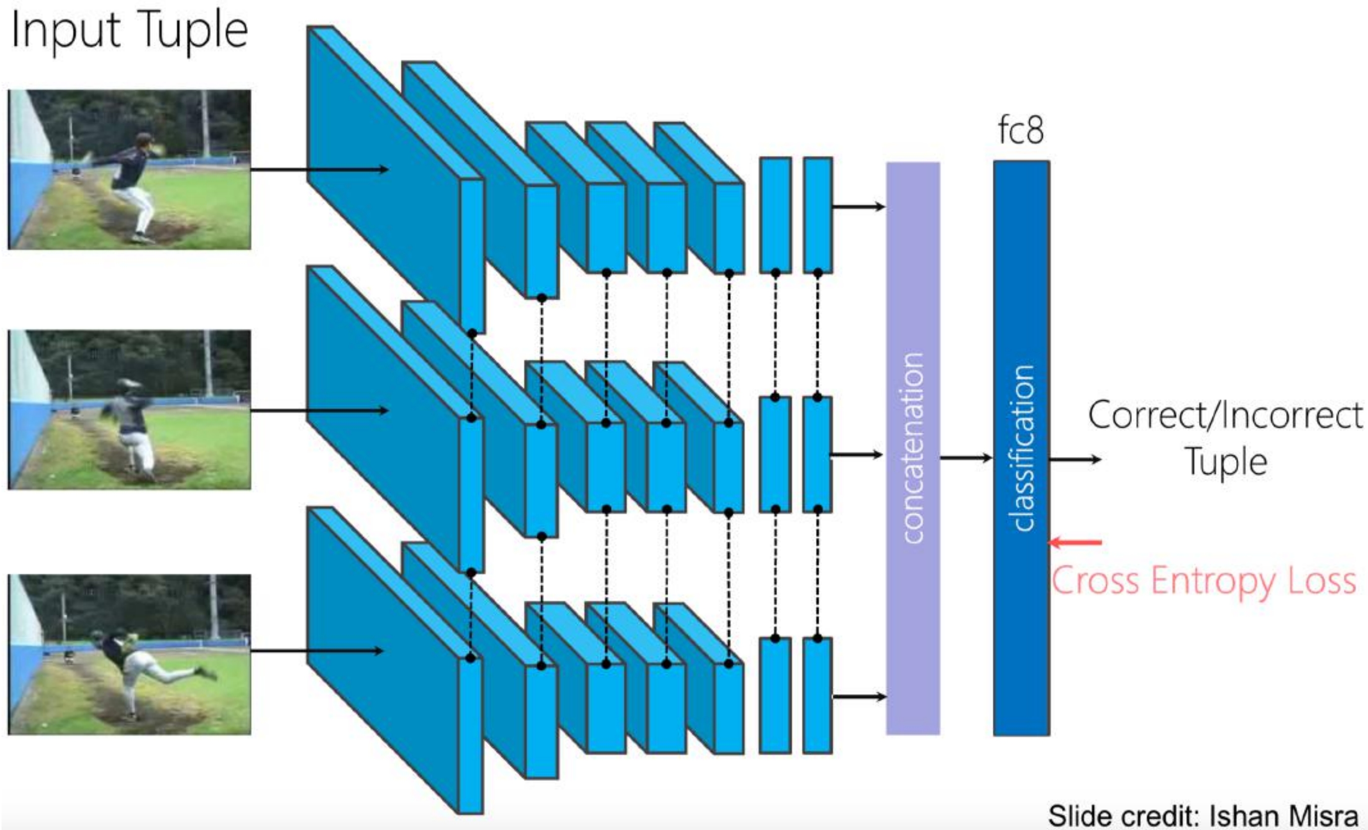
True **False**



Unsupervised Learning using Temporal Order Verification. *Ishan Misra, C. Lawrence Zitnick and Martial Hebert.* ECCV, 2016.

Self-supervised Learning

- Video Example: Shuffle and Learn

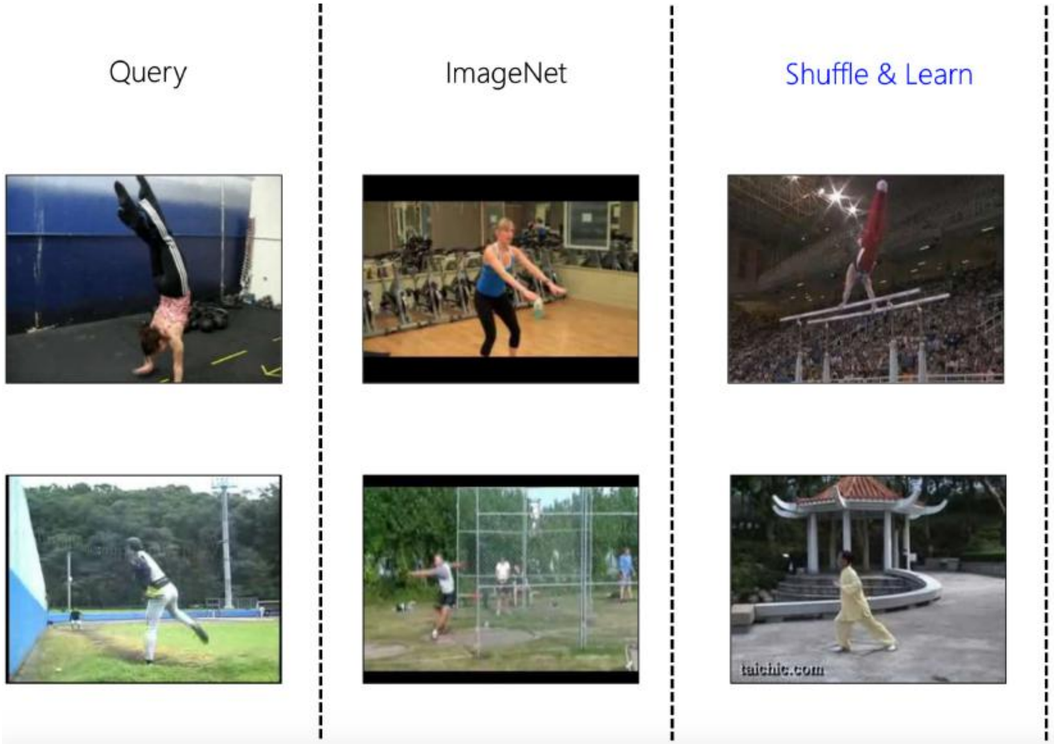


Unsupervised Learning using Temporal Order Verification. *Ishan Misra, C. Lawrence Zitnick and Martial Hebert.* ECCV, 2016.

Self-supervised Learning

- Video Example: Shuffle and Learn

Image Retrieval: Nearest Neighbors of Query Frame (FC5 outputs)

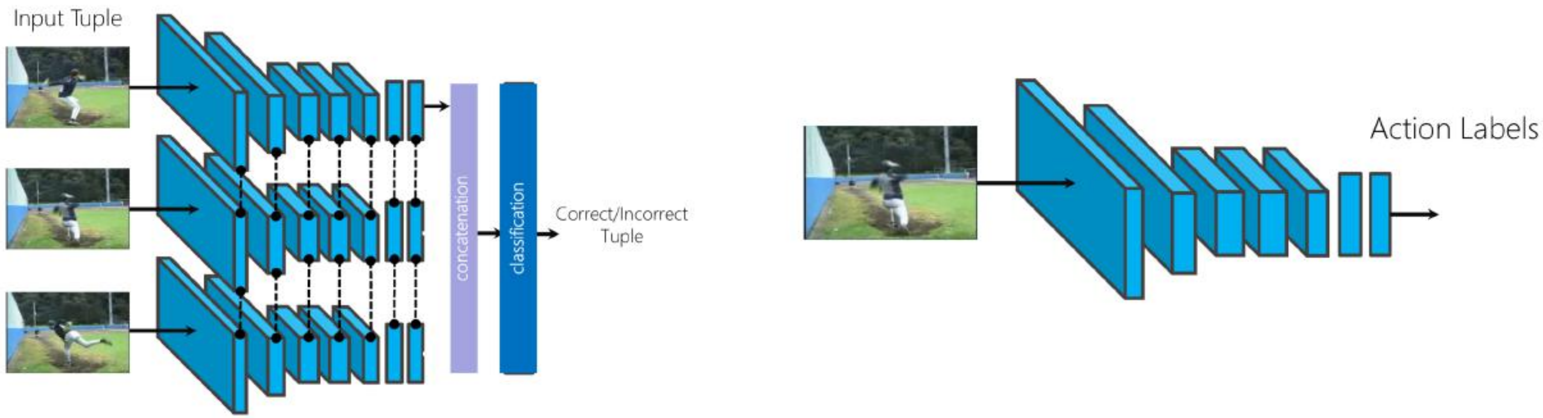


Unsupervised Learning using Temporal Order Verification. *Ishan Misra, C. Lawrence Zitnick and Martial Hebert.* ECCV, 2016.

Self-supervised Learning

- Video Example: Shuffle and Learn

Action Recognition: Self-supervised pre-train + Finetune

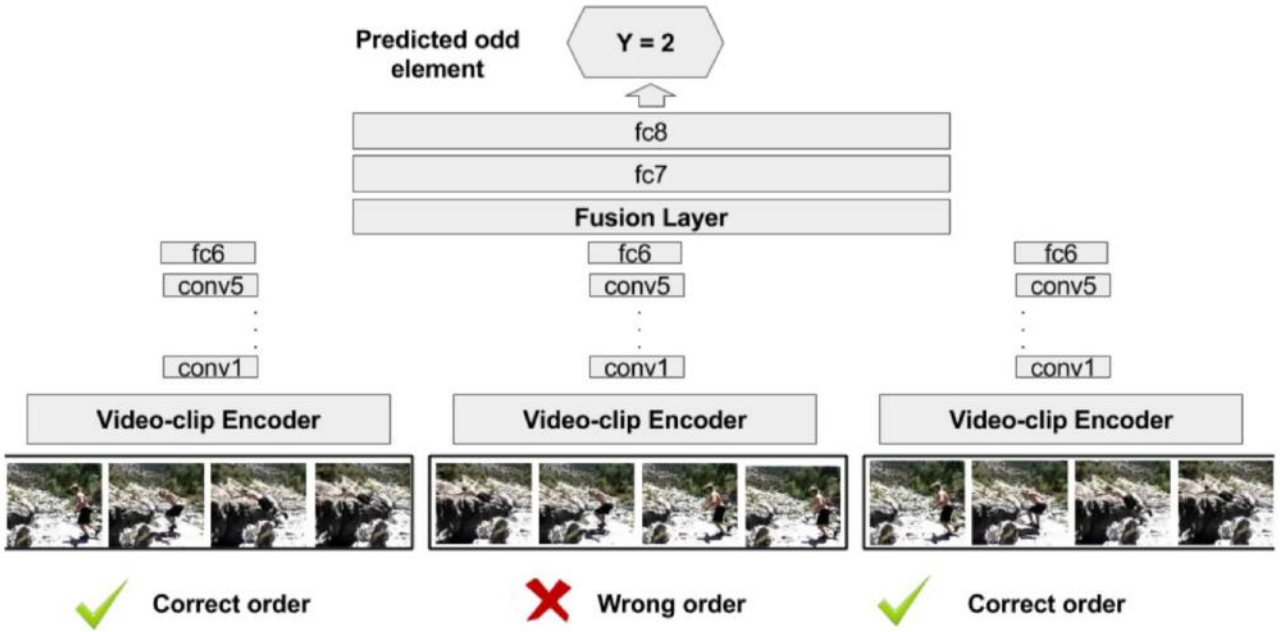


Dataset	Initialization	Mean Classification Accuracy
UCF101	Random	38.6
	Shuffle & Learn	50.2
	ImageNet pre-trained	**67.1**

Unsupervised Learning using Temporal Order Verification. *Ishan Misra, C. Lawrence Zitnick and Martial Hebert.* ECCV, 2016.

Self-supervised Learning

- Video Example: Odd-One-Out

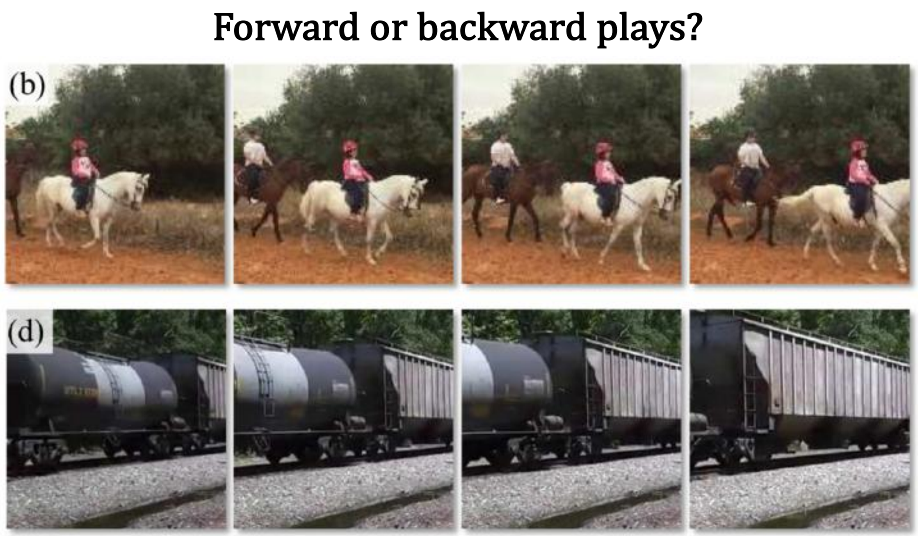


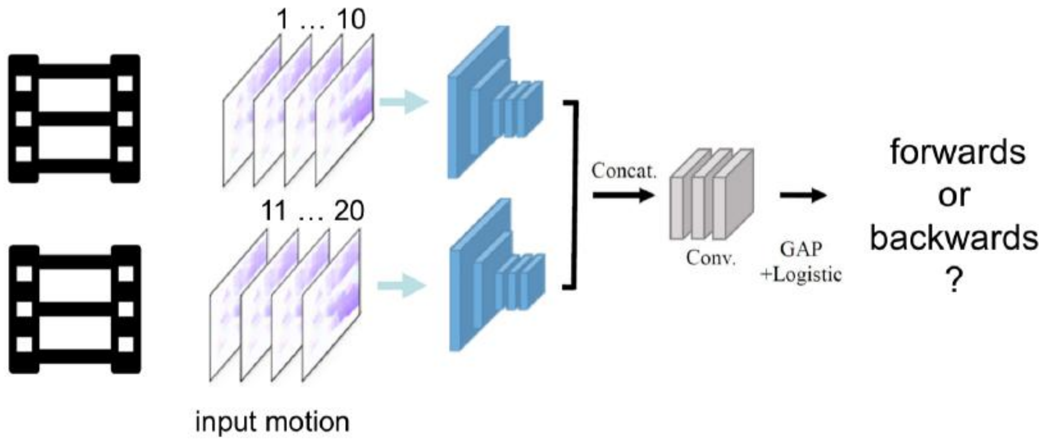
Initialization	Mean Classification Accuracy
Random	38.6
Shuffle and Learn	50.2
Odd-One-Out	60.3
ImageNet pre-trained	**67.1**

Self-Supervised Video Representation Learning With Odd-One-Out Networks. *Basura Fernando, Hakan Bilen, Efstratios Gavves, and Stephen Gould*, ICCV 2017

Self-supervised Learning

- Video Example: Learning the Arrow of Time





- Depending on the video, solving the task may require
(a) low-level understanding (e.g. physics)
(b) high-level reasoning (e.g. semantics)
(c) familiarity with very subtle effects
(d) camera conventions

- Input: optical flow in two chunks
- Final layer: global average pooling to allow class activation map (CAM)

Learning and Using the Arrow of Time. *Donglai Wei, Joseph Lim, Bill Freeman, Andrew Zisserman.* CVPR 2018

Self-supervised Learning

- Video Example: Temporal Coherence of Color

Colorize all frames of a grey scale version using a reference frame

Reference Frame

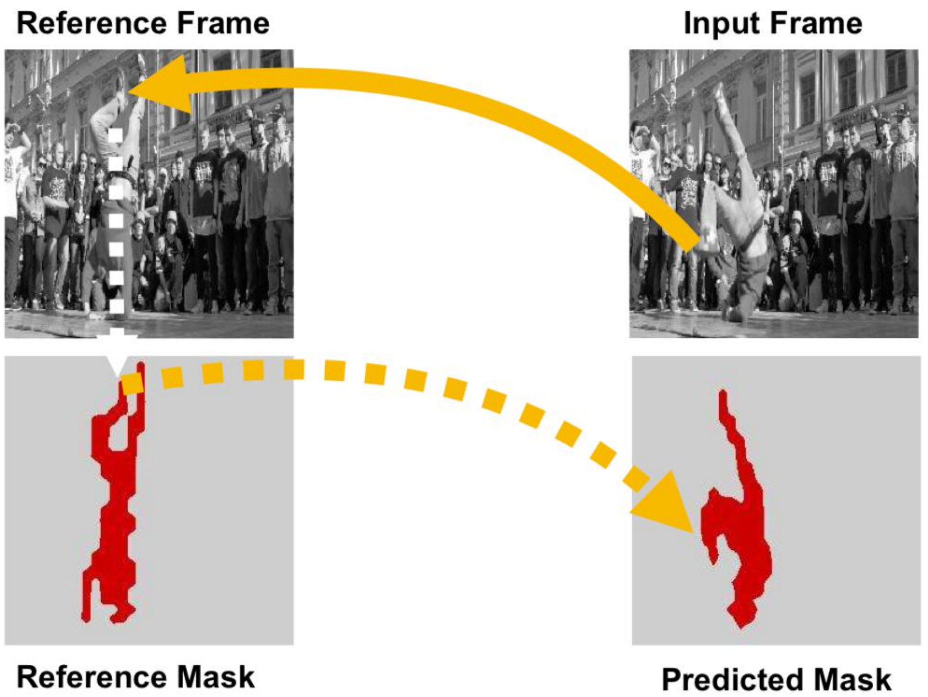
What color is that?

Tracking Emerges by Colorizing Videos
Vondrick, Shrivastava, Fathi, Guadarrama, Murphy, ECCV 2018

Self-supervised Learning

- Video Example: Temporal Coherence of Color

Tracking Emerges: Only the first frame is given, colors indicate different instances



Tracking Emerges by Colorizing Videos
Vondrick, Shrivastava, Fathi, Guadarrama, Murphy, ECCV 2018

Self-supervised Learning

- Video Example: Temporal Coherence of Color

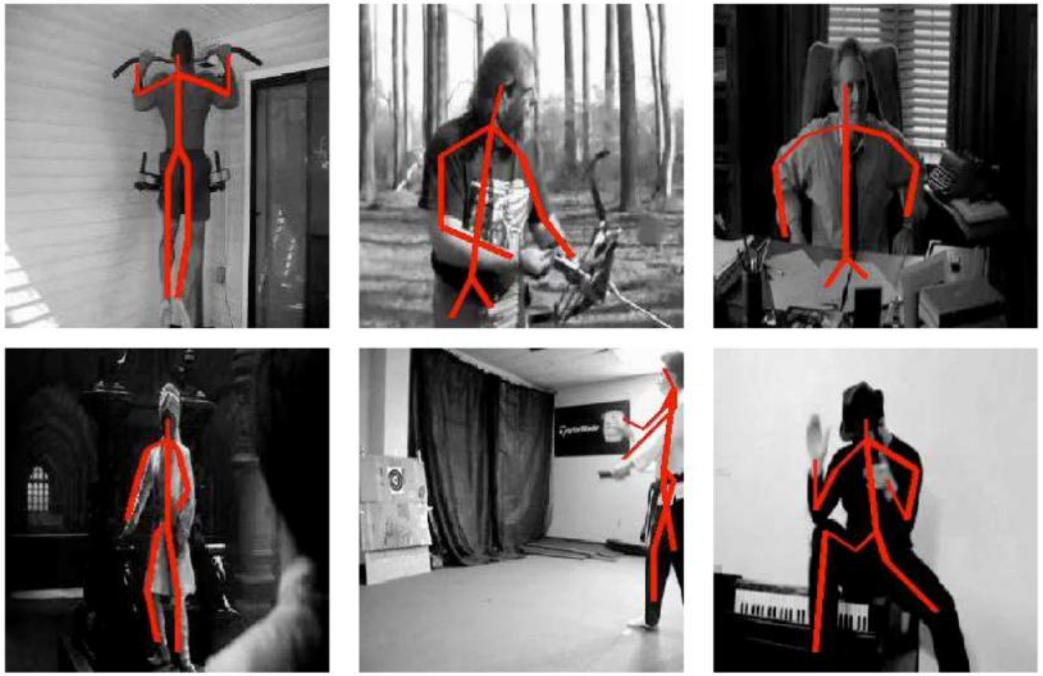
Segment Tracking: Only the first frame is given, colors indicate different instances

Tracking Emerges by Colorizing Videos
Vondrick, Shrivastava, Fathi, Guadarrama, Murphy, ECCV 2018

Self-supervised Learning

- Video Example: Temporal Coherence of Color

Pose Tracking: Only the skeleton in the first frame is given

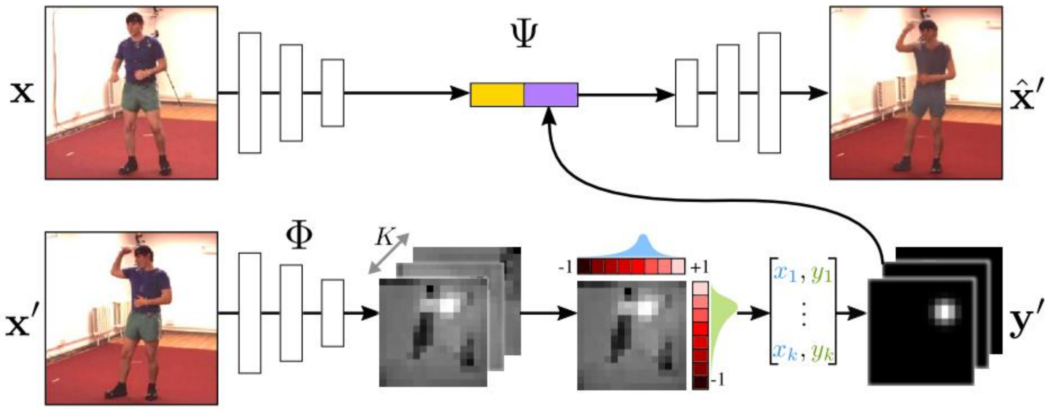


Tracking Emerges by Colorizing Videos
Vondrick, Shrivastava, Fathi, Guadarrama, Murphy, ECCV 2018

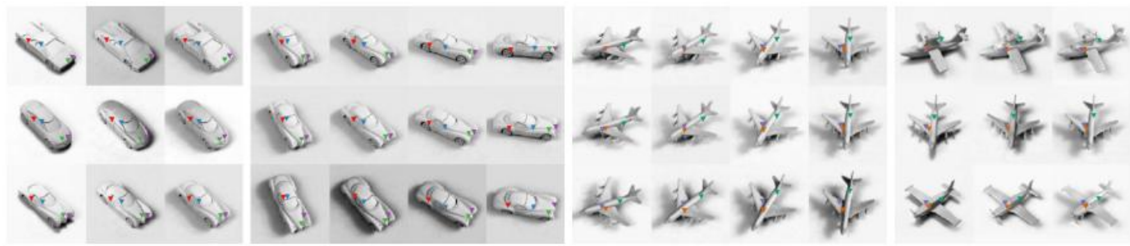
Self-supervised Learning

- Video Example: Temporal Coherence of Color

Unsupervised Key-point Detection: Only paired images of the same object is given



- Achieve retargeting
- Disentangling Style and Geometry
- Invariant Localization



(a) (b) (c) (d)

Unsupervised Learning of Object Landmarks through Conditional Image Generation
Tomas Jakab, Ankush Gupta et al. NIPS, 2018.

Self-supervised Learning

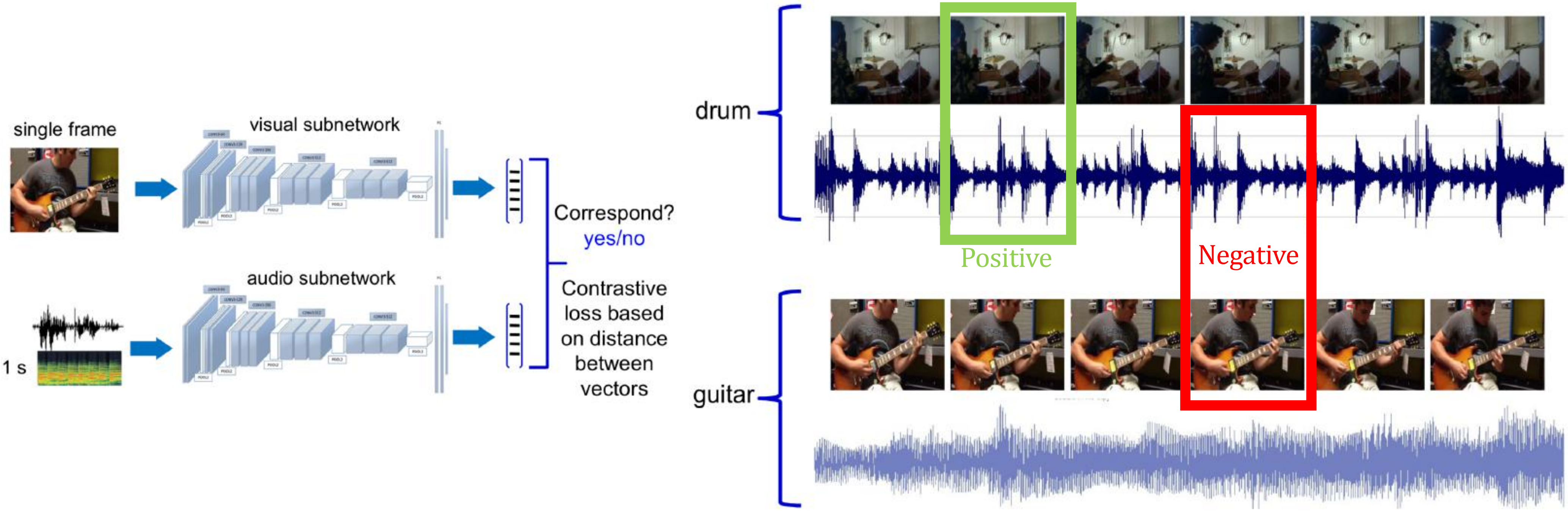
- **Video + Sound Example**

- Sound and frames are:
  - Semantically consistent
  - Synchronized
- Two types of proxy task:
  - Predict audio-visual correspondence
  - Predict audio-visual synchronization

Self-supervised Learning

- Video + Sound Example: Audio-Visual Co-supervision

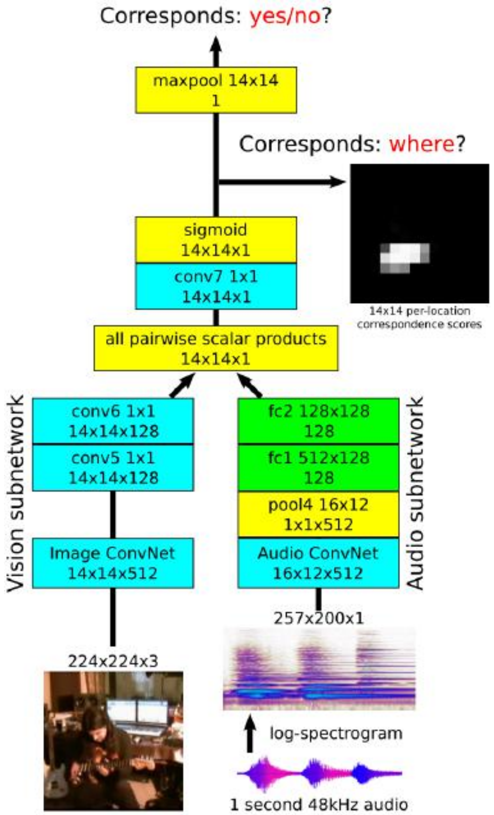
Train a network to predict if image and audio clip correspond



Objects that Sound. *Arandjelović and Zisserman*, ICCV 2017 & ECCV 2018

Self-supervised Learning

- Video + Sound Example: Audio-Visual Co-supervision



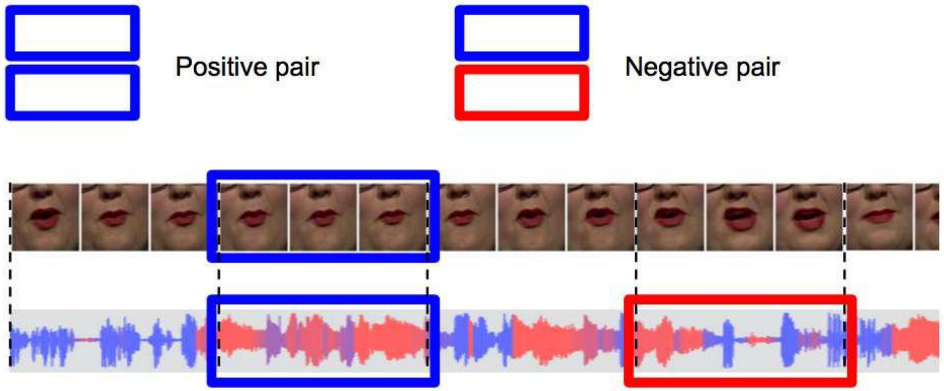
- Learn good visual features
- Learn good audio features
- Learn aligned audio-visual embeddings
- Learn to localize objects that sound

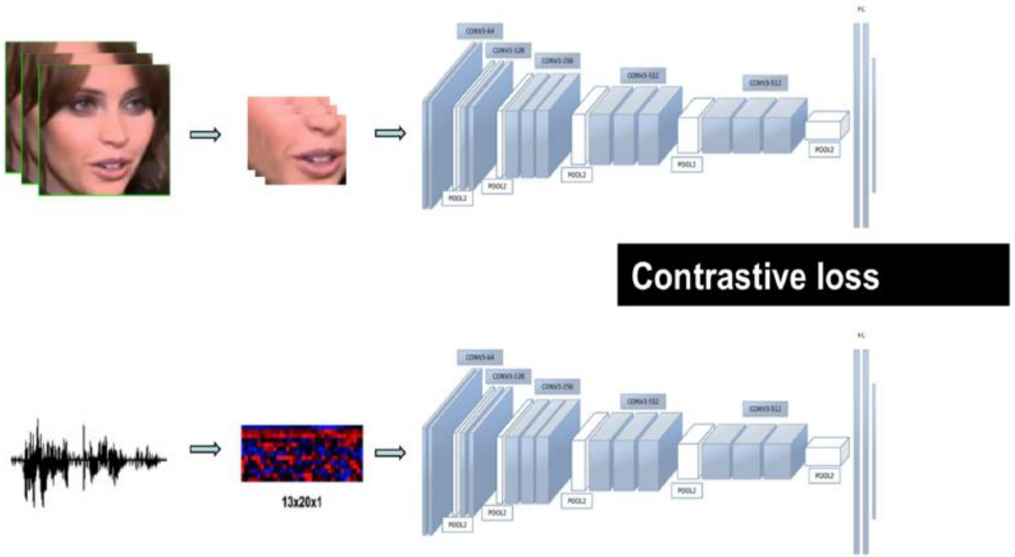
- Using learned features
  - Sound classification
  - Query on image to retrieve audio
  - Localizing objects with sound

Objects that Sound. *Arandjelović and Zisserman* *(DeepMind, Ox)*, ICCV 2017 & ECCV 2018

Self-supervised Learning

- Video + Sound Example: Audio-Visual Co-supervision





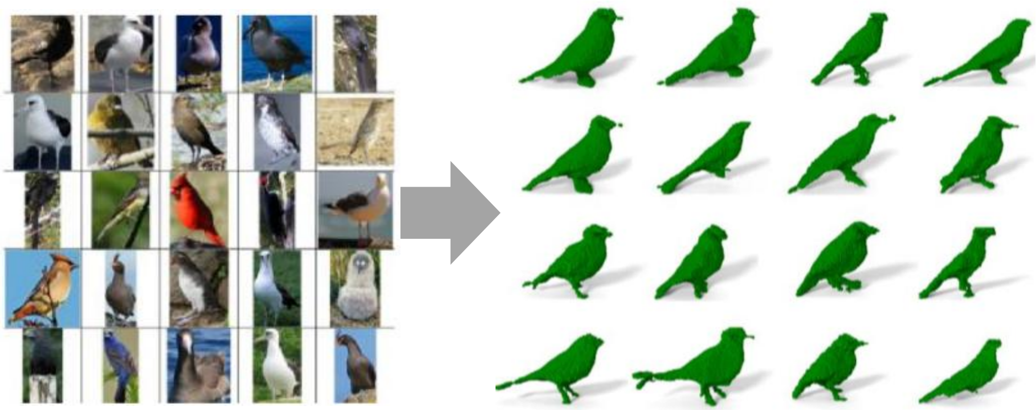
- Applications
  - Active speaker detection
  - Audio-to-video synchronization
  - Voice-over rejection
  - Visual features for lip reading

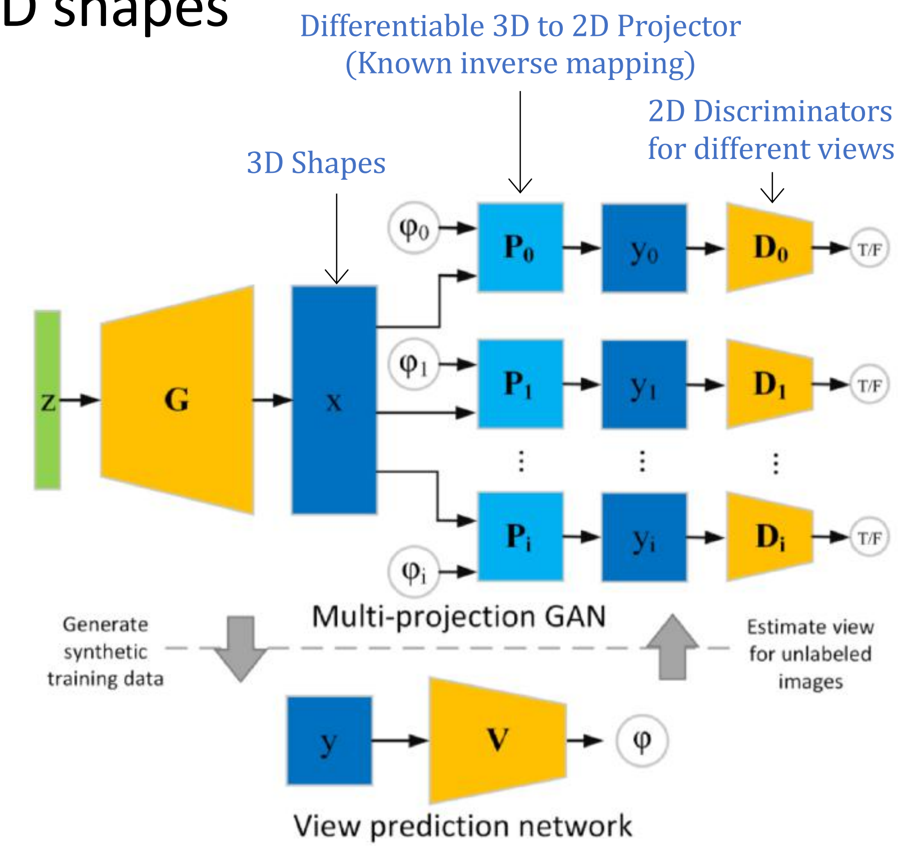
Out of time: Automatic lip sync in the wild. *Chung, Zisserman,* 2016

Self-augmented Learning

Self-augmented Learning

- Example: Unsupervised 2D images to 3D shapes

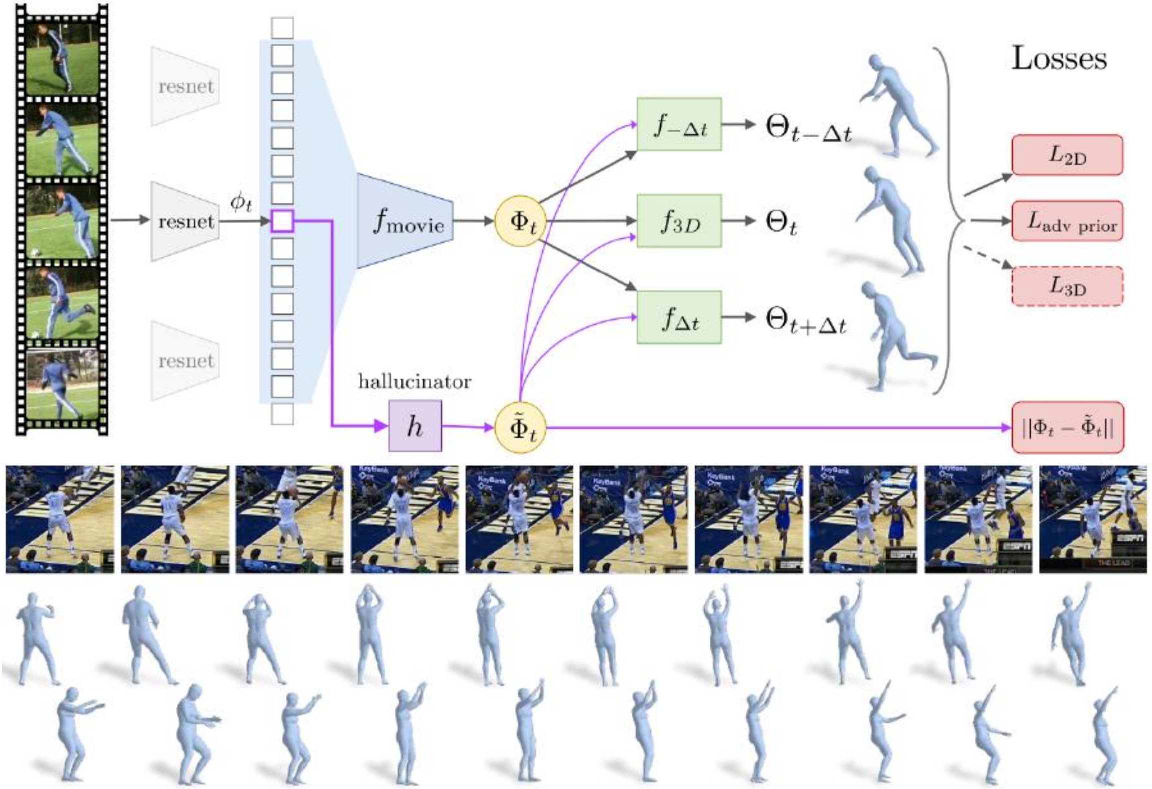




Synthesizing 3D Shapes from Unannotated Image Collections using Multi-projection Generative Adversarial Networks.
Xiao Li, Yue Dong, Pieter Peers, Xin Tong. CVPR, 2019

Self-augmented Learning

- Example: 2D Video to 3D shape



Learning 3D Human Dynamics from Video*. A. Kanazawa, J. Zhang et al*. CVPR, 2019

Summary

Dual, Self-Supervised, Self-augmented Learnings

- Dual, Self-supervised, Self-augmented Learnings
- Dual Learning
- Self-supervised Learning
- Self-augmented Learning

Dual, Self-Supervised, Self-augmented Learnings

- References
  - Dual Learning: A New Learning Paradigm
    https://www.youtube.com/watch?v=HzokNo3g63E
  - DeepMind: Self-supervised Learning
    https://project.inria.fr/paiss/files/2018/07/zisserman-self-supervised.pdf
  - Learning Discrete Representations via Information Maximizing Self-Augmented Training http://proceedings.mlr.press/v70/hu17b/hu17b.pdf

Dual, Self-Supervised, Self-augmented Learnings

- Exercise 1: (Optional)
  - Choice an application and implement it

Link: https://github.com/zsdonghao/deep-learning-note/

北京大學
PEKING UNIVERSITY

Questions?