Research Seminar

Y-Data Deep Learning Course

Goals

Reading scientific papers

Presenting the new concept of a paper

Engage with the paper in a deep level
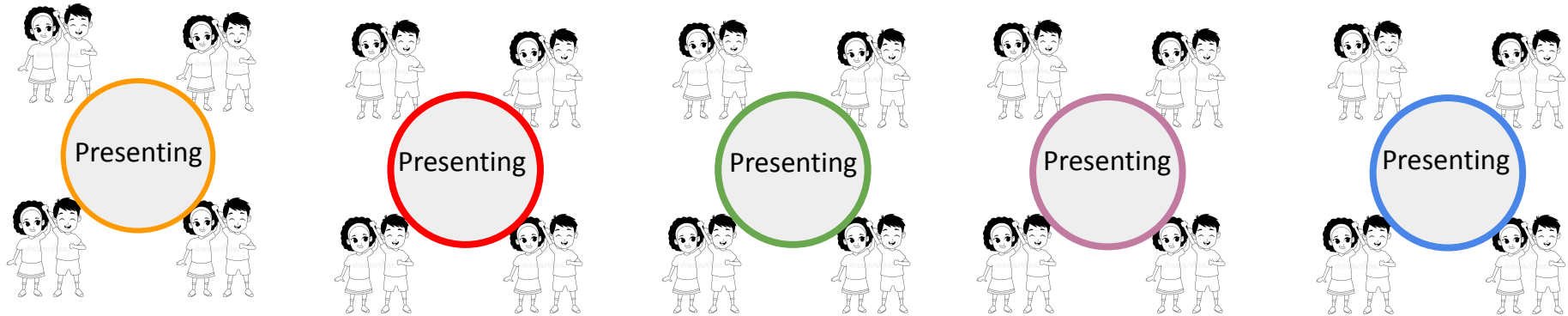
[Research seminar - student presentation]

25.3
29.4
17.6

Be familiar with state-of-the-art research in the field beyond what is taught in class

[Research seminar - paper presentation]

11.3
6.5
10.6

Research seminar sessions

5 groups on 8-10 students (4-5 pairs of students)

Presenting   Presenting   Presenting   Presenting   Presenting

Assignments:

Presenting the paper	~30 min
Task coding (1) on the paper	~20 min
Task coding (2) on the paper	~20 min
Presenting the previous work\be a reviewer	~20 min

You should send the slides to the mentor a day before the presentation. Every group should prepare presentation

Colorization as a Proxy Task for Visual Understanding
Larsson et al. 2017

Research Seminar
29.4

Col

isual

Colorization as a Proxy Task for Visual Understanding

Gustav Larsson
University of Chicago
larsson@cs.uchicago.edu

Michael Maire
TTI Chicago
mmaire@ttic.edu

Gregory Shakhnarovich
TTI Chicago
greg@ttic.edu

Abstract

We investigate and improve self-supervision as a drop-in replacement for ImageNet pretraining, focusing on automatic colorization as the proxy task. Self-supervised training has been shown to be more promising for utilizing unlabeled data than other, traditional unsupervised learning methods. We build on this success and evaluate the ability of our self-supervised network in several contexts. On VOC segmentation and classification tasks, we present results that are state-of-the-art among methods not using ImageNet labels for pretraining representations.

Moreover, we present the first in-depth analysis of self-supervision via colorization, concluding that formulation of the loss, training details and network architecture play important roles in its effectiveness. This investigation is further expanded by revisiting the ImageNet pretraining paradigm, asking questions such as: How much training data is needed? How many labels are needed? How much do features change when fine-tuned? We relate these questions back to self-supervision by showing that colorization provides a similarly powerful supervisory signal as various flavors of ImageNet pretraining.

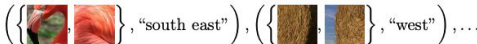**Learning a representation via (x, y) pairs**

Classification

(, "flamingo") , (, "hay") , ...

Self-supervision

Ex. 1: **Inpainting** (remove patch and then predict it)

(,) , (,) , ...

Ex. 2: **Context** (given two patches, predict their spatial relation)

({ , } , "south east") , ({ , } , "west") , ...

Ex. 3: **Colorization** (predict color given intensity)
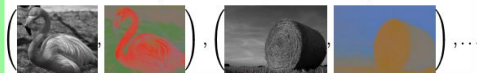
(,) , (,) , ...

Figure 1. Using a representation that was originally trained for classification on (x, y) pairs to initialize a network has become standard practice in computer vision. Self-supervision is a family of alternative pretraining methods that do not require any labeled data, since labels are "manufactured" through unlabeled data. We

Research seminar - 29.4

Self-supervised learning (SSL)

unsupervised   supervised

self-supervised

Research seminar - 29.4

Self-supervised learning (SSL)

<u>unsupervised</u>

Proxy task is solved based on pseudo-labels which help to initialize the network weights.

<u>supervised</u>

Actual task is performed (downstream task)

Research seminar - 29.4

Self-supervised learning (SSL)

Why?

For transfer learning - dataset with small number of labels

<u>unsupervised</u>

Proxy task is solved based on pseudo-labels which help to initialize the network weights.
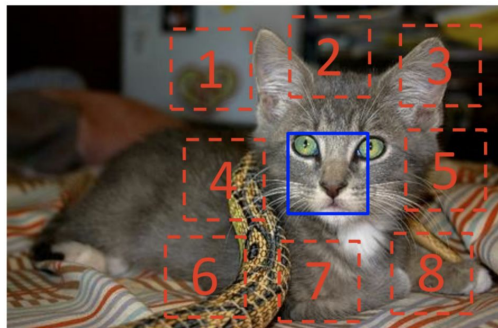
<u>supervised</u>

Actual task is performed (downstream task)

Research seminar - 29.4

Self-supervised learning (SSL)

 Why?

 For transfer learning - dataset with small number of labels

unsupervised

Proxy task is solved based on pseudo-labels which help to initialize the network weights.

supervised

Actual task is performed (downstream task)
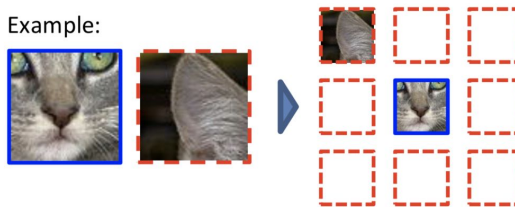
Classification task

Research seminar - 29.4

Self-supervised learning (SSL)

Proxy task - solving puzzle

$X = ($ � , ◀ $); Y = 3$

Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. Noroozi and Favaro. 2017.

Research seminar - 29.4

Self-supervised learning (SSL)

Proxy task - predict rotations

Objectives:

| | Rotate 0 degrees | Rotated image: X^0 | ConvNet model F(.) | Maximize prob. $F^0(X^0)$ | Predict 0 degrees rotation (y=0) |

$g(X, y=0)$ — Rotate 0 degrees — Rotated image: X^0 — ConvNet model F(.) — Maximize prob. $F^0(X^0)$ — Predict 0 degrees rotation (y=0)

$g(X, y=1)$ — Rotate 90 degrees — Rotated image: X^1 — ConvNet model F(.) — Maximize prob. $F^1(X^1)$ — Predict 90 degrees rotation (y=1)

$g(X, y=2)$ — Rotate 180 degrees — Rotated image: X^2 — ConvNet model F(.) — Maximize prob. $F^2(X^2)$ — Predict 180 degrees rotation (y=2)

$g(X, y=3)$ — Rotate 270 degrees — Rotated image: X^3 — ConvNet model F(.) — Maximize prob. $F^3(X^3)$ — Predict 270 degrees rotation (y=3)

Image X

UNSUPERVISED REPRESENTATION LEARNING BY PREDICTING IMAGE ROTATIONS. Gidaris et al. 2018

Research seminar - 29.4

Self-supervised learning (SSL)

Paper proxy task - colorization
Predict the color components from the intensity component

Ex. 3: **Colorization** (predict color given intensity)

Research seminar - 29.4

Assignments:

All students must read the paper

Team 1: Presenting the paper - **30 minutes** (do not cover previous work)
- You should present the idea of the paper, the training details, results and evaluations
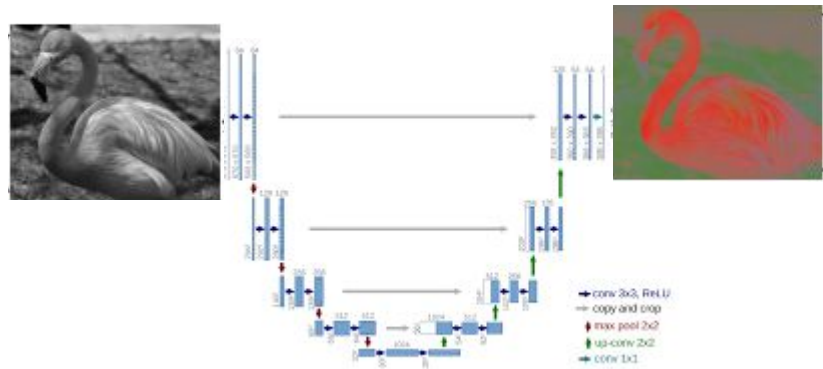
Self supervised - 5 min
Main idea - 5 min
Training details - 5 min
Experiments - 10 min

Research seminar - 29.4

Assignments:

All students must read the paper

Team 2: Coding. Implement the main idea of the paper - **20 minutes**:

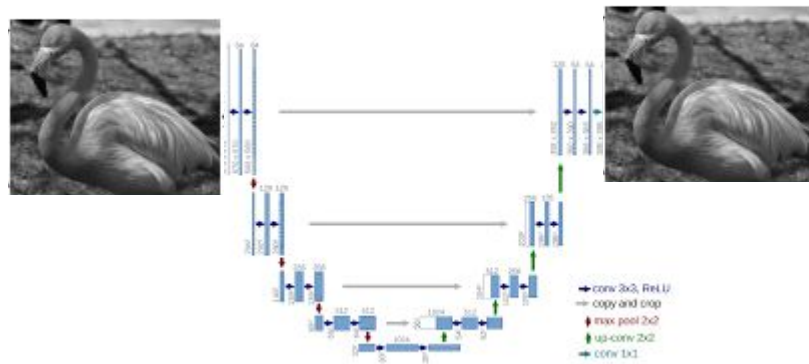Colorization proxy task -> downstream (target) task

- Better to use the architecture and the dataset of the paper - but not mandatory. You can use whatever network and dataset you want.
- Downstream task on politicians classification task

Code is written in caffe - not recommended

Research seminar - 29.4

Assignments:

All students must read the paper

Team 3: Coding. Implement the main idea of the paper - **20 minutes**:

Denoising auto-encoder -> downstream (target) task

- Better to use the architecture and the dataset of the paper - but not mandatory. You can use whatever network and dataset you want.
- Downstream task on politicians classification task

Code is written in caffe - not recommended

Research seminar - 29.4

Assignments:

All students must read the paper

Team 4: Be a reviewer - **20 minutes**

a) Elaborate on another self-supervised task (can be in text, images… but not what already introduced in class)

b) What kind of papers cited this paper? (introduce one of them shortly)

c) Do you have any criticism on the paper?

Another self supervised paper (just mention the idea) - 10 min
One paper that cited this paper - 5 min
Other criticism you have from reading the paper? Are the experiments enough? (in comparison to other papers in this topic) - 5 min.