DAAD 16th Workshop "Software Engineering Education and Reverse Engineering"

COMPUTER VISION APPROACHES BASED ON DEEP LEARNING AND NEURAL NETWORKS

A Systematic Mapping Study

Eralda Nishani, Prof.Dr. Betim Çiço
South East European University, Tetovo, Macedonia

Overview

- Introduction to Deep NN

PART I

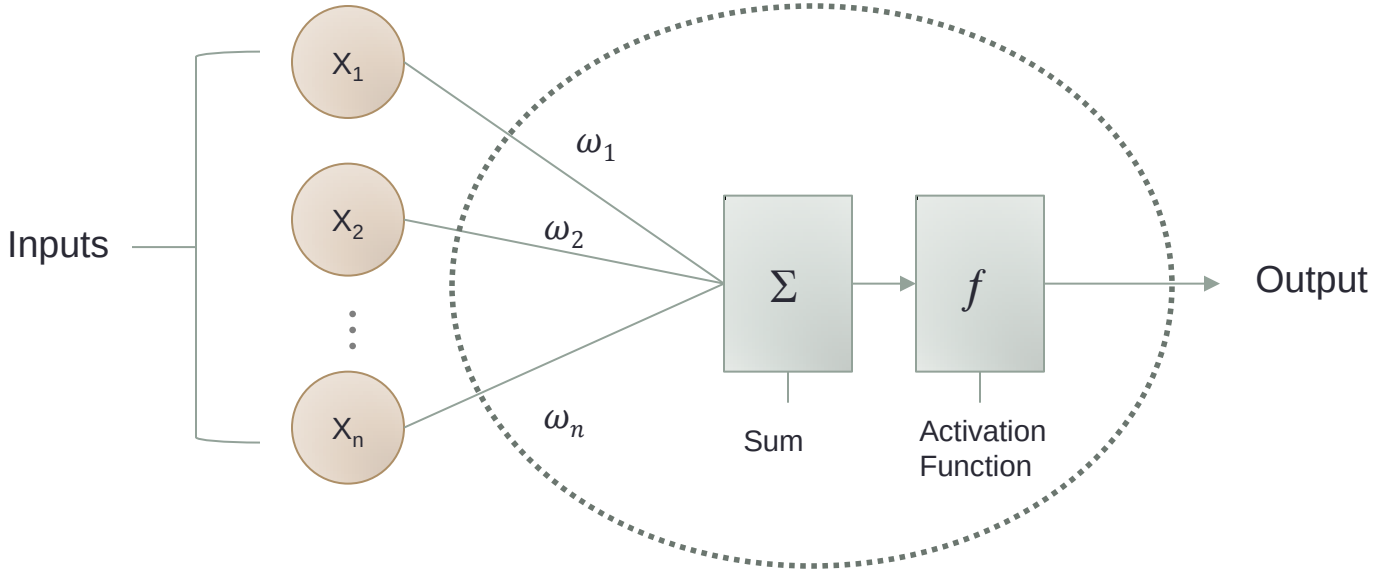
- Systematic mapping study
  - Research questions
  - Classification scheme
- Results from research questions

PART II

- Possible research area – Deep NN for Human Pose Estimation

Deep Learning in Neural Networks

- "*A class of machine learning techniques, where many hierarchical layers of information-processing stages are exploited for pattern classification and for feature or representation learning.*" *[Li Deng, 2015]*
- Artificial Neural Network (NN) consists of multiple connected neurons
- In order to make a NN exhibit certain behavior deep learning is necessary



Deep NN and Graphics Processing Units

- The popularity of deep learning has increased
  - Increased chip processing possibilities
  - Lowered cost of computer processing hardware
  - Large amount of data existing online
- (GP)GPUs are gaining more usage for
  - Training deep NN with larger datasets
  - Accelerating the training process

Deep NN and Computer Vision

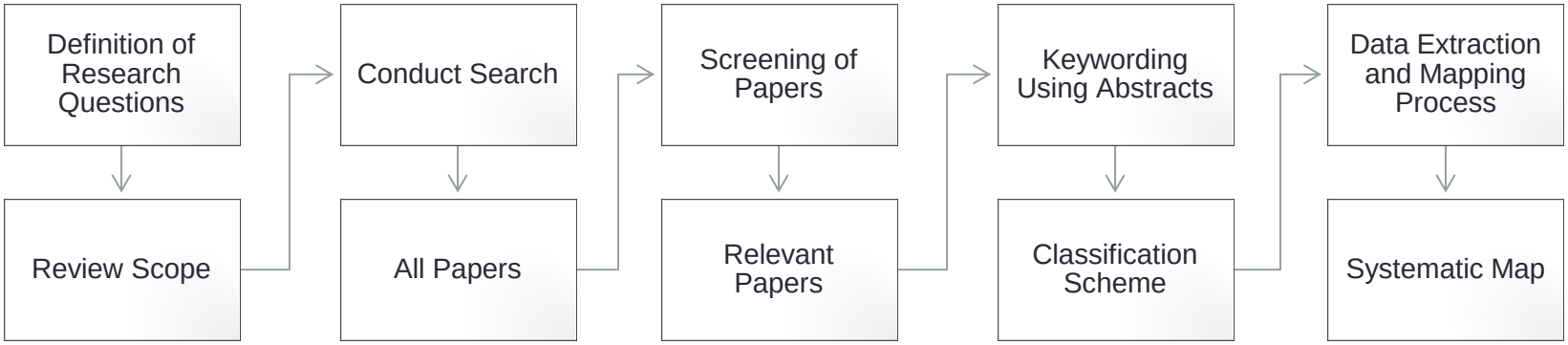
- Computer vision – finding features from images/videos to help discriminate objects
- The beginning – human-driven approach
  - Specific algorithms – formulate a way to represent the image by encoding features
- The next step – Deep NN
  - For each layer the image is transformed and abstract features are discovered
  - These features serve for different tasks (detection, classification, localization, segmentation)
  - Convolutional Neural Networks - *[Krizhevsky et al, 2012]:*

    "AlexNet" – proved that CNN was suitable for computer vision tasks

PART I

Systematic Mapping Study

Systematic Mapping Study



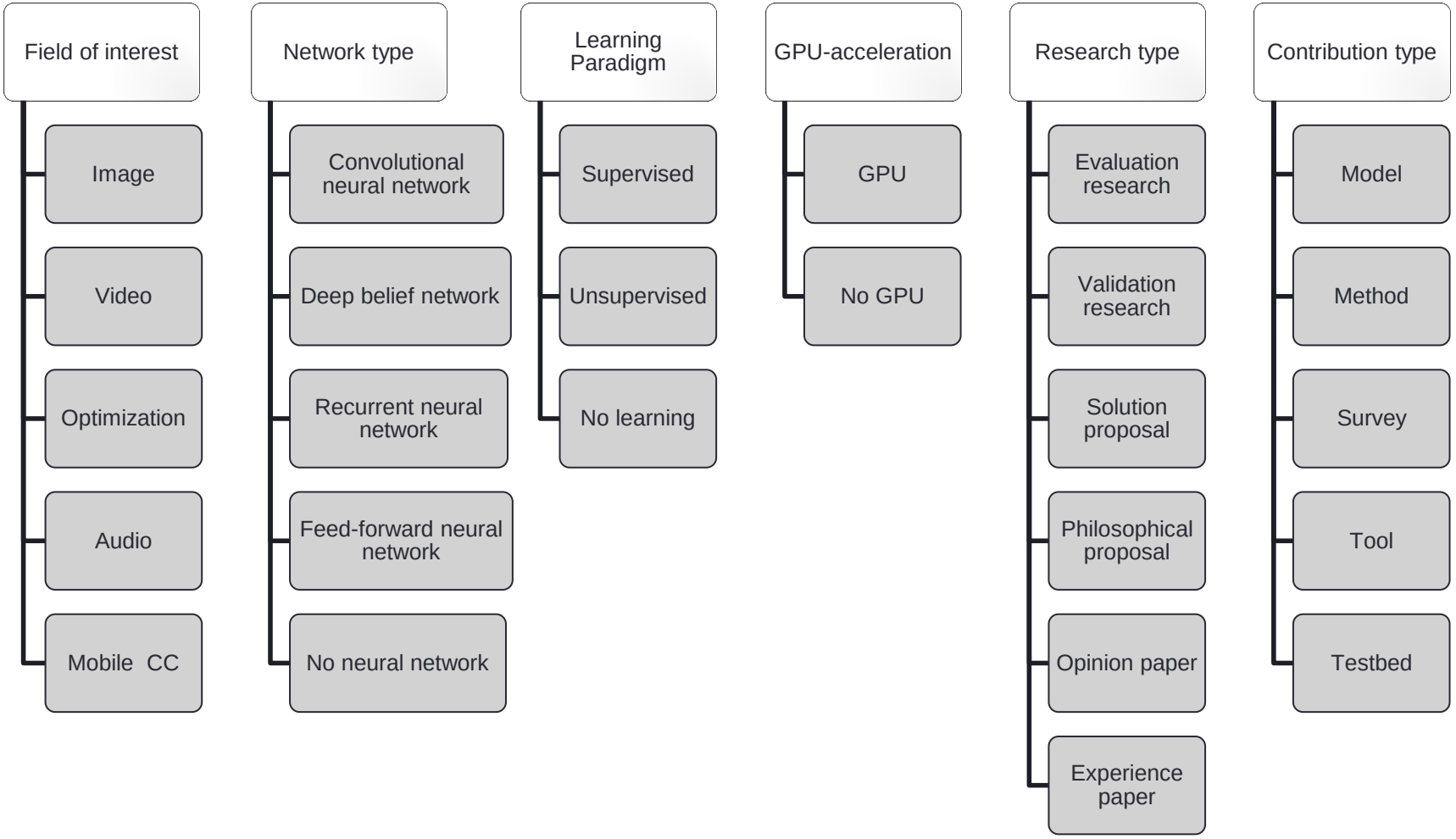
No.	Research Question	Motivation
RQ1	What is the main field of interest investigated in the papers?	This question defines the main goal of a paper and helps find out where recent studies have been focusing on.
RQ2	What is the type of neural network architecture used?	This question provides information over the architecture used to implement a solution.
RQ3	What type of learning is used to train the neural network?	This question is used to find our whether the neural networks are trained or not.
RQ4	Is the implementation GPU-accelerated?	This question is related to the fact whether or not an implementation includes GPU.
RQ5	What kinds of research and contributions have been proposed?	This question provides information about the concrete techniques and solutions that have been proposed.
RQ6	How publications have evolved over time?	This question shows the evolution of the publications concerning the subject under study.

Systematic Mapping Study

- Resource – IEEE Xplore Digital Library
- From 263 articles ⟹ 119 final articles

No.	Search String	No. of papers
SS1	((("Abstract":deep learning algorithm) OR "Abstract":convolutional network) OR "Abstract":GPU)	20
SS2	((("deep learning algorithm") AND "convolutional network") OR GPU)	74
SS3	(((((("deep learning algorithm") OR "convolutional network") AND GPU) AND image) AND video)	42
SS4	((((("deep learning algorithm") OR "convolutional network") AND image) AND video)	127

Classification scheme



Research Questions

- RQ1: What are the main field of interest investigated in the papers?
- RQ2: What kind of neural network are mentioned in the papers?

Field of interest	Number of papers	Percentage
Image	75	63.03%
Video	24	20.17%
Optimization	14	11.76%
Audio	4	3.36%
Mobile CC	2	1.68%

Network Type	Number of papers	Percentage
Convolutional neural network	78	65.55%
Feed-forward neural network	29	24.37%
Deep belief network	5	4.20%
Recurrent neural network	3	2.52%
No neural network	4	3.36%

Research Questions

- RQ3: What kind of learning paradigms are mentioned in the papers?
- RQ4: Is GPU mentioned for acceleration in the selected papers?

Learning paradigm	Number of of papers	Percentage
Supervised	65	54.62%
Unsupervised	32	26.89%
No learning	22	18.49%

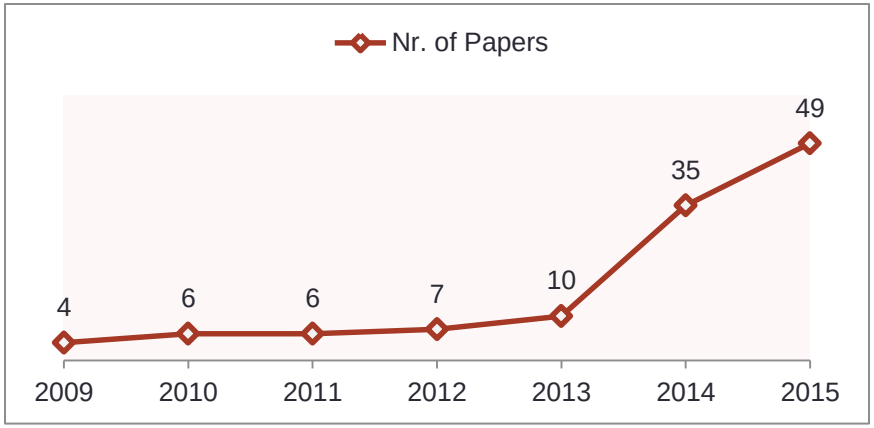
GPU-acceleration	Number of papers	Percentage
GPU	71	59.66%
No GPU	48	40.34%

Research Questions

- RQ5: What kinds of research and contributions have been proposed?
- RQ6: How have publications changed over time?

Research Type	Number of papers	Percentage
Evaluation research	51	42.86%
Validation research	33	27.73%
Solution proposal	22	18.49%
Philosophical proposal	7	5.88%
Opinion paper	5	4.20%
Experience paper	1	0.84%

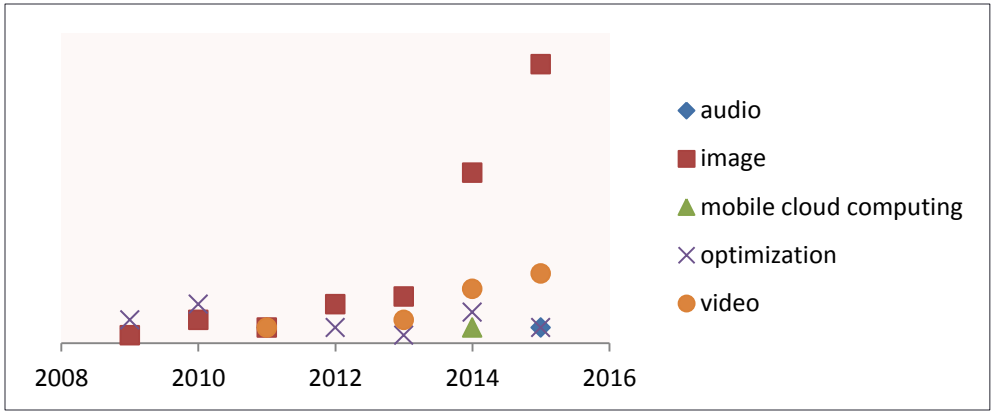
Contribution Type	Number of papers	Percentage
Model	63	52.94%
Method	47	39.50%
Survey	5	4.20%
Tool	3	2.52%
Testbed	1	0.84%



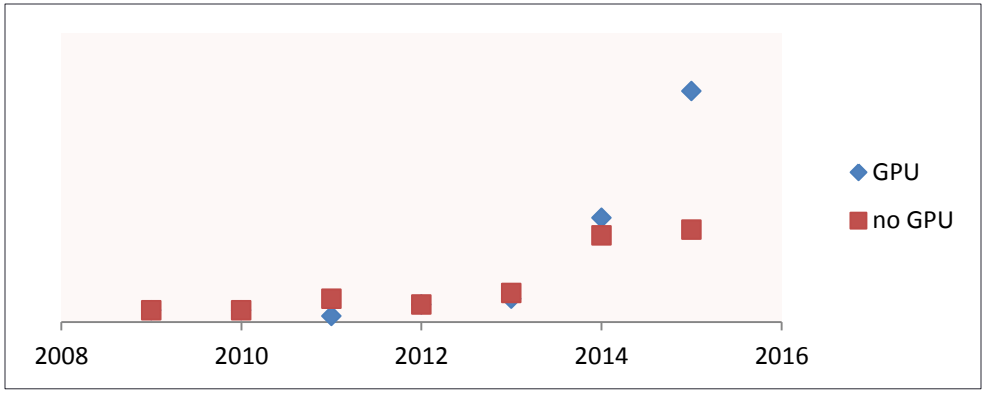
Discussion - Time series

Recently:

- Focus has shifted on image search approaches



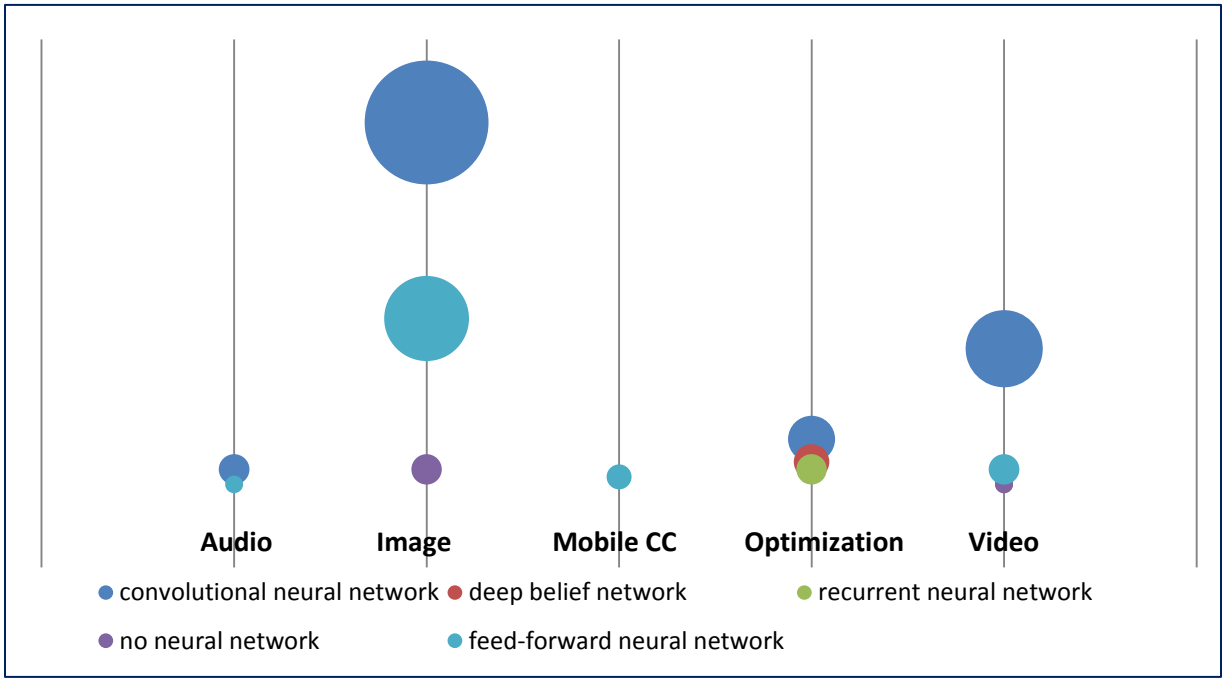
- GPU-acceleration is becoming popular

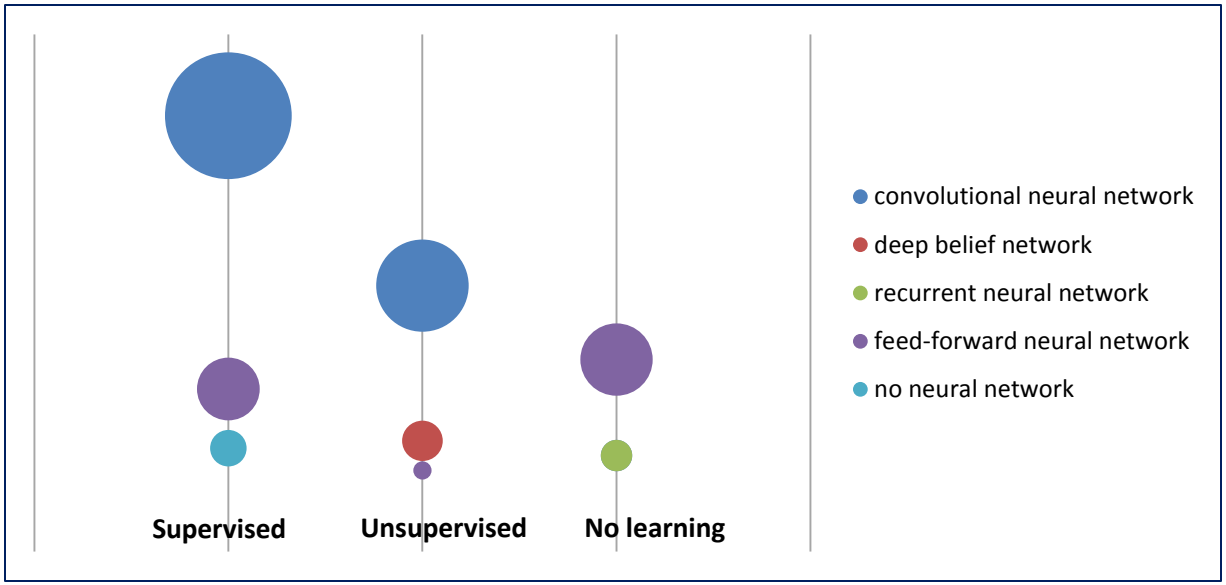


Discussion - Relationships

- For image-search approaches convolutional networks are preferred

- Supervised learning is used for convolutional neural networks





Results of Study

- The majority of papers:
  - Deal with image search approaches (63.03 %)
  - Use convolutional neural networks (65.55 %)
  - Use supervised learning paradigm (54.62 %)
  - Make use of GPU-acceleration (59.66 %)
- We can conclude that this field is relevant to the researchers community

PART II

Possible research area – Deep NN for Human Pose Estimation

Human Pose Estimation

- Objective – track the position of human body parts in images and videos
- Applications
  - Human-computer interaction
  - Gaming
  - Gesture recognition
- Challenges
  - Variable human appearance in images
  - Variable human body shape
  - Motion blur
  - Occlusions of human body parts

Overview

- Earliest methods
  - Place as pattern recognition problem
  - Learn a ‘body plan’ using statistical techniques
- Pictorial structures
  - Model the body parts as a Conditional Random Field (CRF)
  - An object is a collection of parts arranged in deformable configuration
- Random Forests
  - Holistic approach
  - Class of machine learning methods
- Deep convolutional networks

State-of-the-art

- DeepPose *[Toshev & Szegedy,2014]*
  - Formulate the task as a joint regression problem
  - Holistic method
- Multi-source deep model *[Ouyang et al., 2014]*
- R-CNN - CNN trained jointly for multiple tasks: pose estimation and action classification *[Gkioxari et al., 2014]*
- CoordinateNet
  - Coordinates of the joints treated as regression problem
  - Network trained on multiple frames *[Pfister, 2015]*
- HeatmapNet
  - Heatmap of positions are the target of regression *[Pfister, 2015]*

Human pose estimation in video frames

1. CNN have been proved successful for human pose estimation task. How can we improve their architecture?
   - *Use a combination of CNN flavors for different sub-tasks*

2. Based on the temporal information provided by video frames, are Recurrent Neural Networks useful for this task?

3. Can we use unsupervised learning to leverage the huge amount of unlabeled data that exists online?
   - *The fact that videos have spatial and temporal regularities can help to build unsupervised learning models*

THANK YOU!

QUESTIONS?