Course: Deep Learning

Unit 2: Computer Vision

Representation Learning

Luis Baumela

Universidad Politécnica de Madrid

Representation Learning

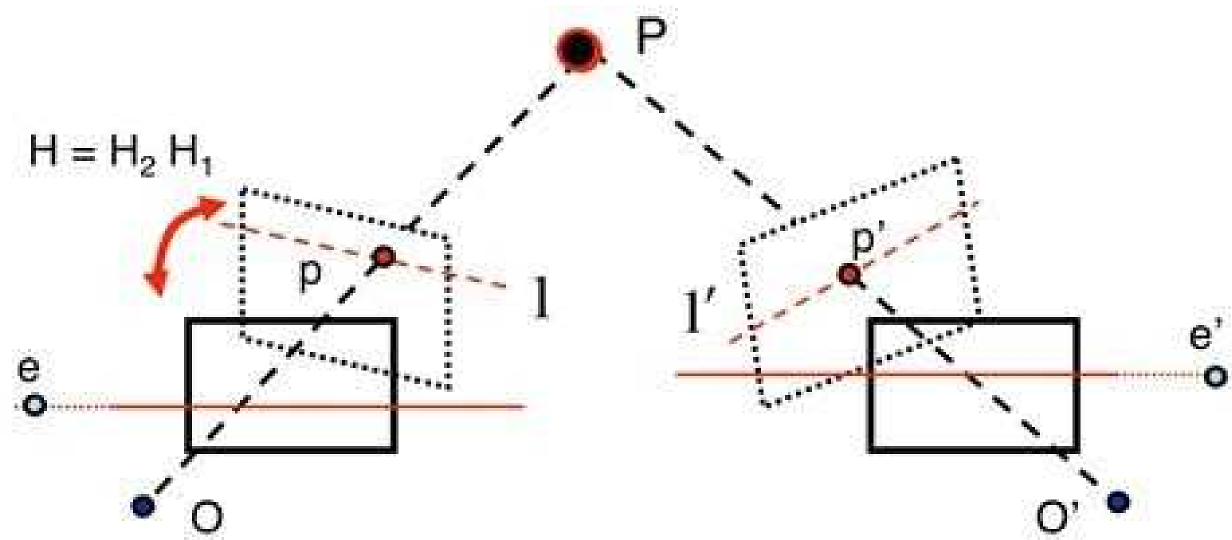
1. Introduction
    - What is a representation?
    - Why are hierarchical representations so good for AI?
2. Learning strategies
    - Transfer learning
    - Multi-task learning
    - Self-supervised learning
    - Semi-supervised + self-supervised learning
    - Contrastive learning
    - Domain adaptation
3. Conclusion

Deep hierarchical representations

- What is a representation?

A formal system that
• makes explicit certain entities and types of information,
• can be operated on by an algorithm
to achieve some information processing goal.

Representations differ in terms of what information they make explicit and of what algorithms they support.



Deep hierarchical representations

- What is a representation?

A formal system that
• makes explicit certain entities and types of information,
• can be operated on by an algorithm
to achieve some information processing goal.

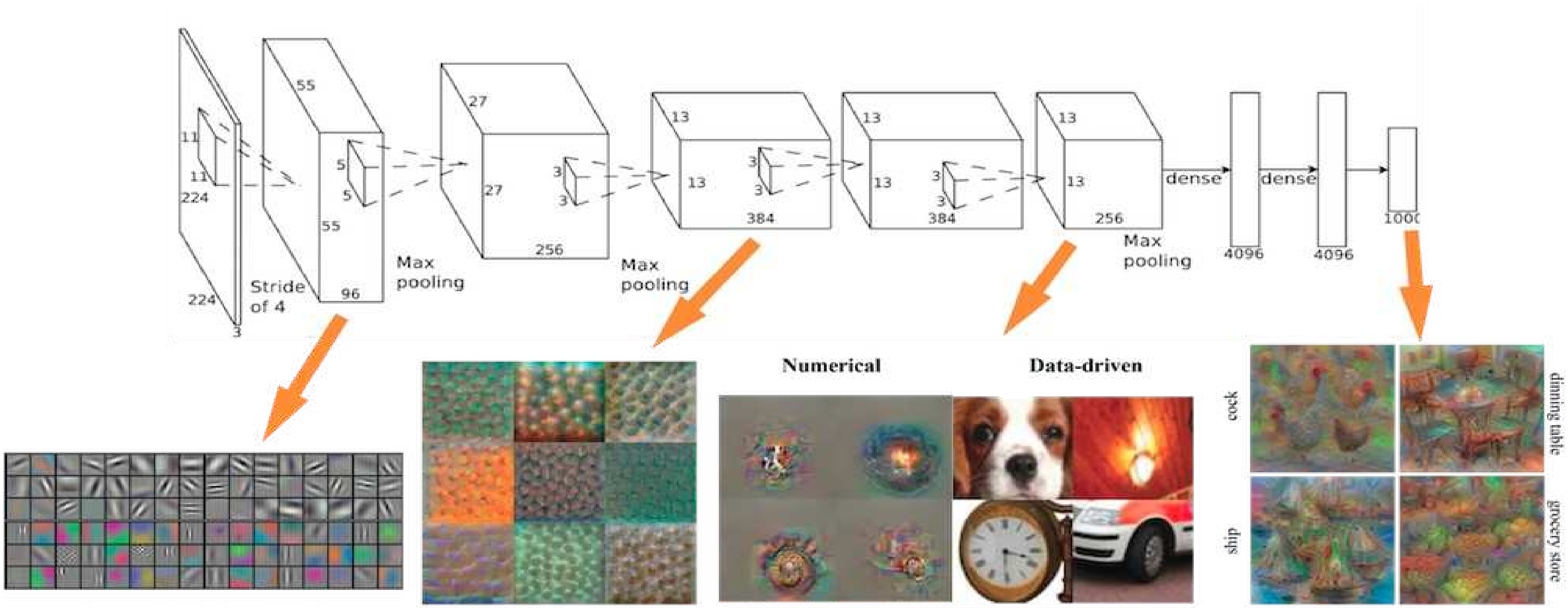
Representations differ in terms of what information they make explicit and in terms of what algorithms they support.

A a deep neural net is a representation of a problem that has emerged from the process of training.

Deep hierarchical representations

- Deep learning is about

  Deep Learning is related to learning **hierarchical representations**, organized in multiple layers.



Conv 1: Edge+Blob
Conv 3: Texture
Conv 5: Object Parts
Fc8: Object Classes

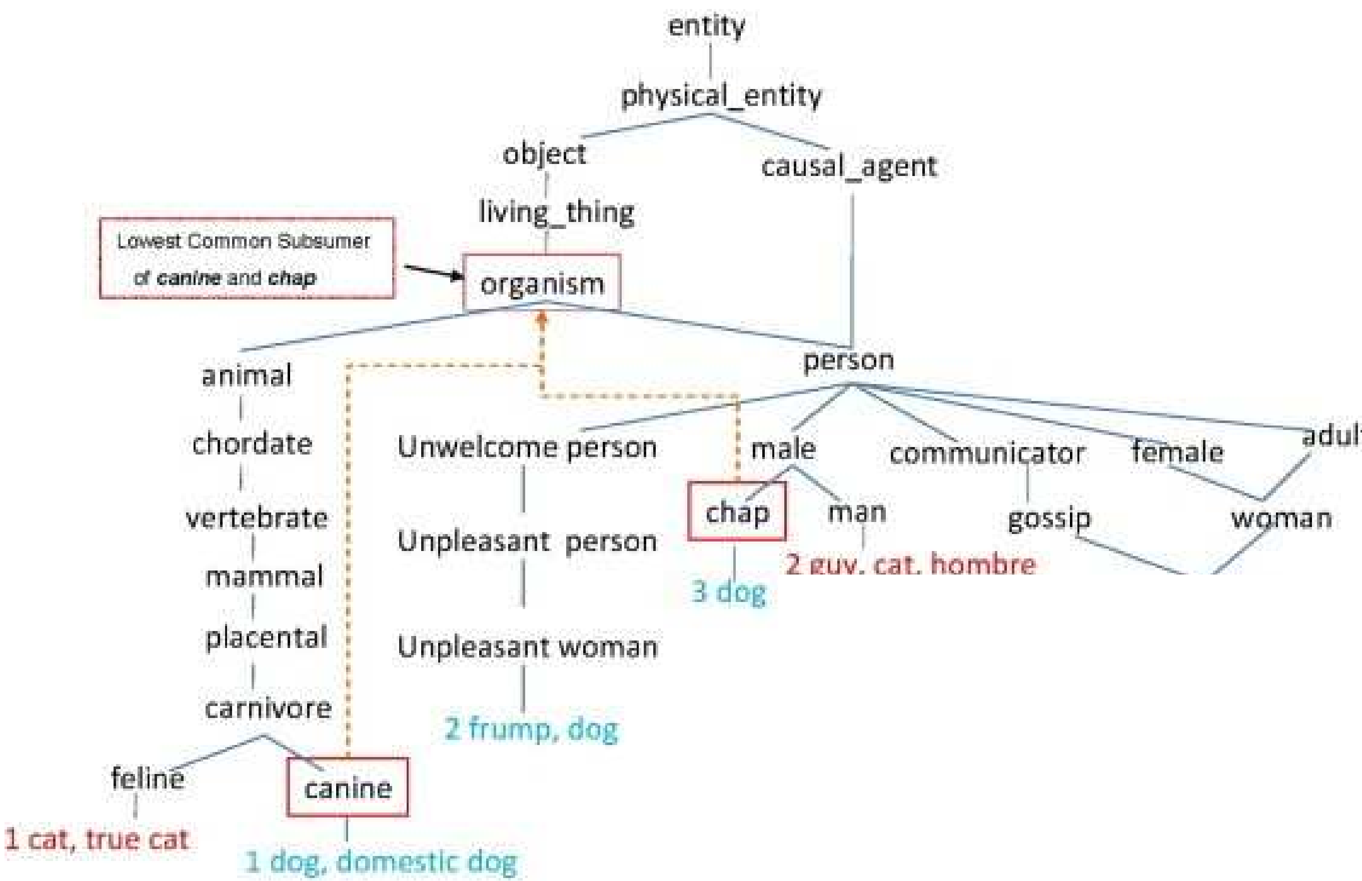
https://wordnet.princeton.edu/

Deep hierarchical representations

- Deep learning is about

  Deep Learning is related to learning **hierarchical representations**, organized in multiple layers.

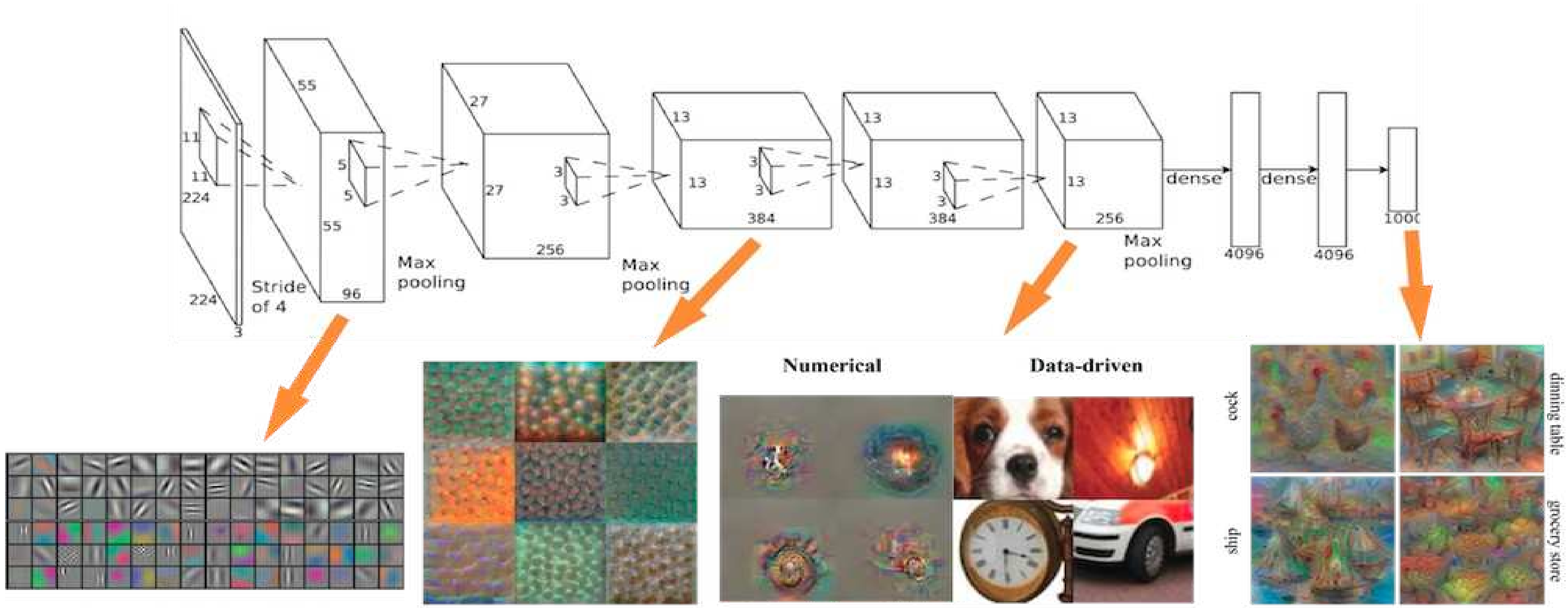
  This hierarchical representation is related to a **hierarchy of abstractions,** like the one we use in language.



https://wordnet.princeton.edu/

Deep hierarchical representations

- Advantage of distributed/hierarchical representation



Conv 1: Edge+Blob | Conv 3: Texture | Conv 5: Object Parts | Fc8: Object Classes

A Deep Neural net can discover features independently of each other that better generalize to unseen samples, hence requiring less training data.

Deep hierarchical representations

- Deep modes for AI

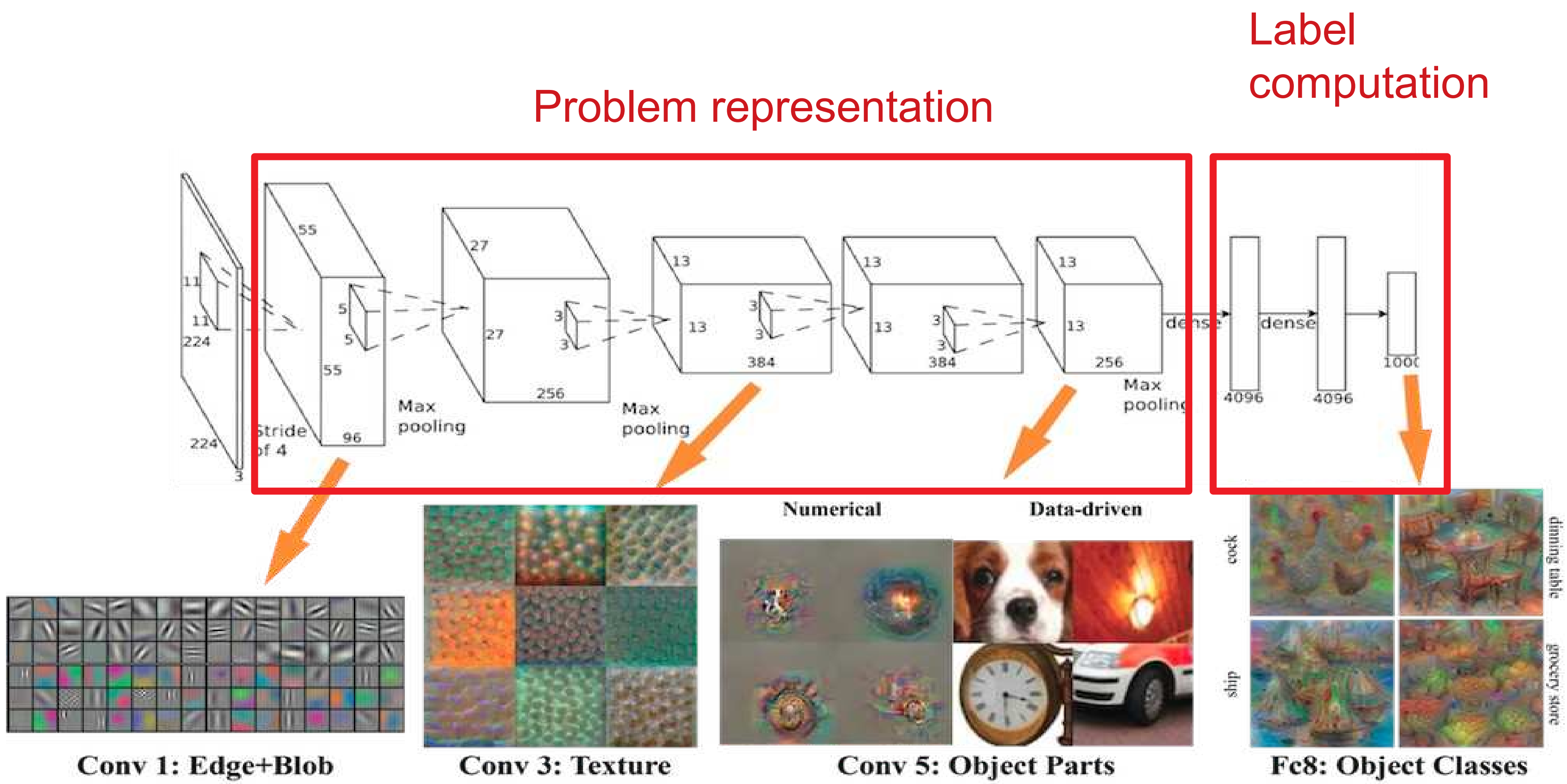
  Building models of the world around us
  - complex enough to represent real world situations,
  - with large (but always limited) amounts of data.

  A hierarchical representation seems adequate to represent AI problems.

  Deep models (e.g. CNNs) provide two new key ingredients to buid AI model:
  - Hierarchical representation
  - Powerful regularization mechanism

Learning strategies

- Data efficiency. Addressing the lack of data.



A deep model provides:

- Top performing solution.
- Hierarchical representation.

The problem representation:

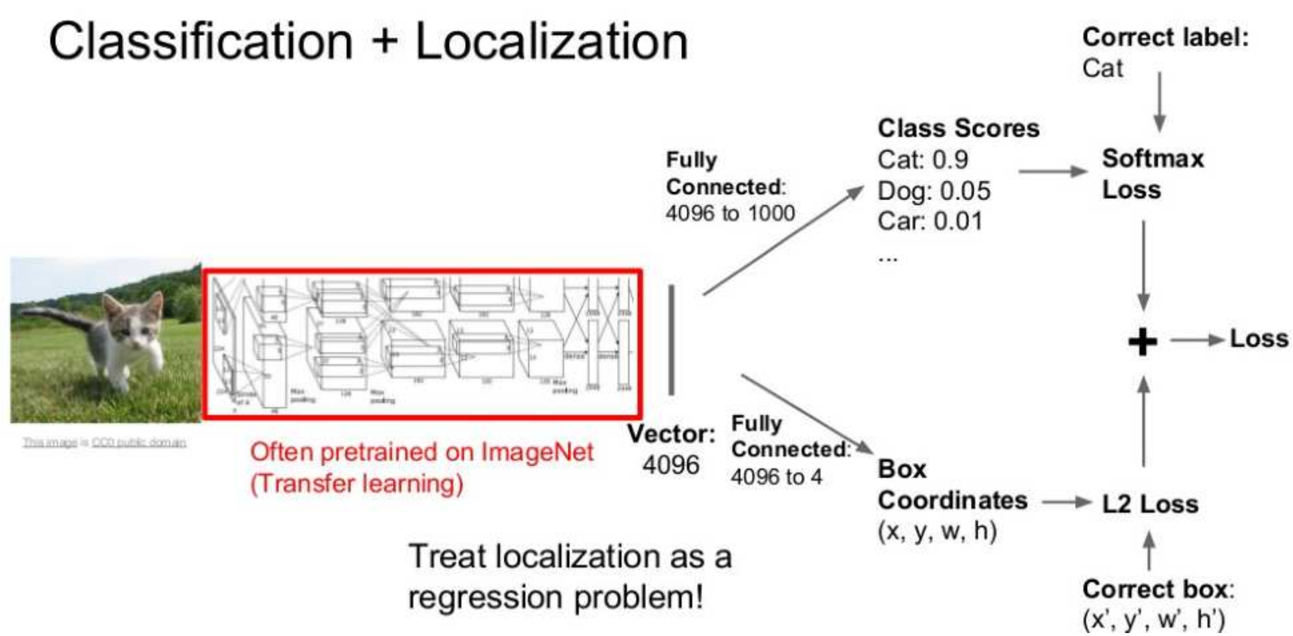
- Is the key for the performance.
- May be reused for other tasks.

Learning strategies

- What is a machine learning task?

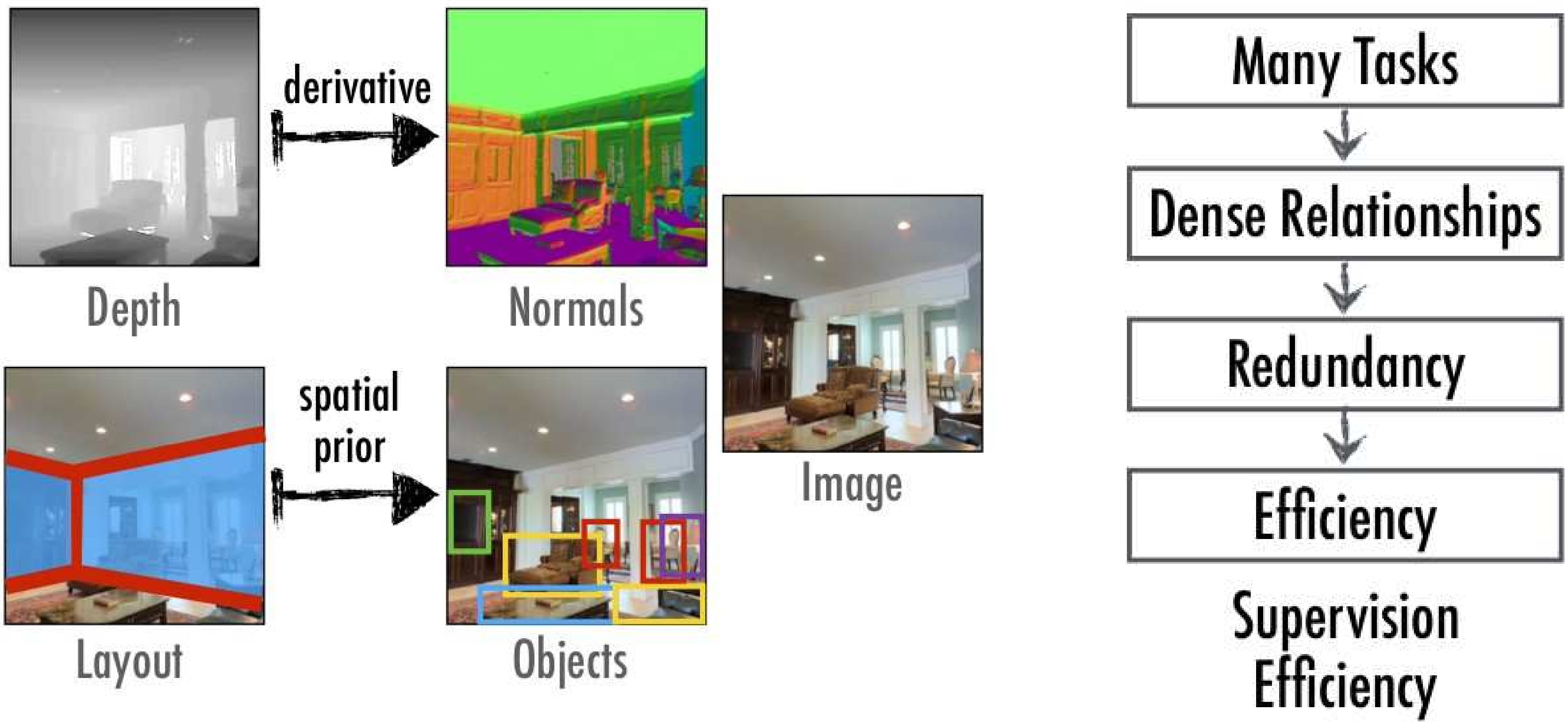
  A problem to be solved by a computational model, with
  - an associated data set of examples; and
  - a function representing the desired solution.

  For example, object classification / localization



Learning strategies

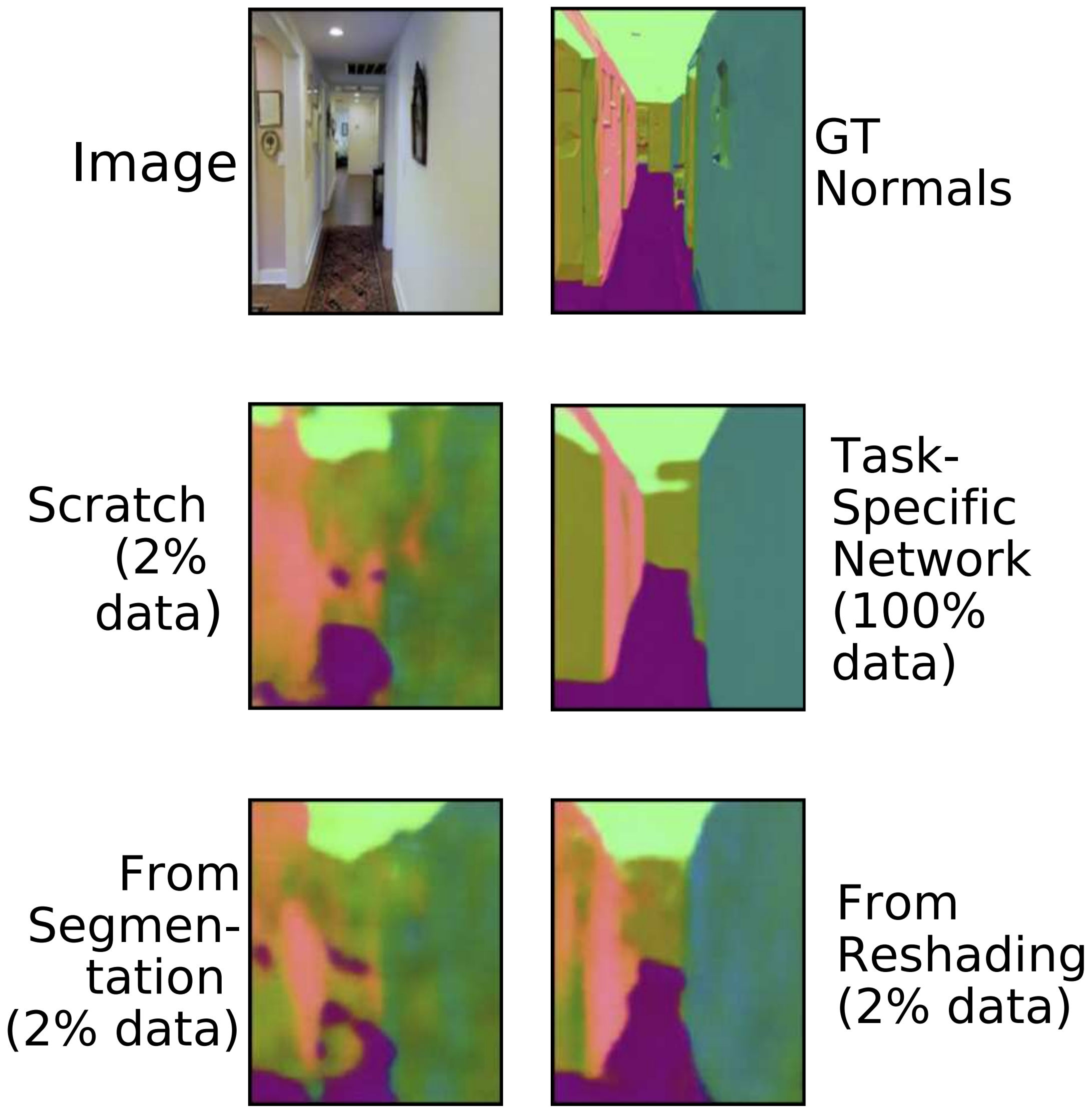
- Are computer vision problems/tasks related?

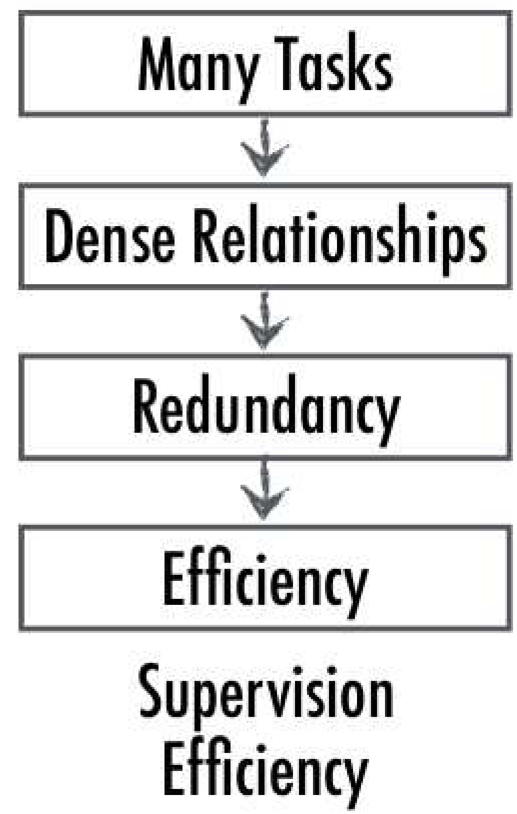


Zamir, Taskonomy: Disentangling Task Transfer Learning CVPR 2018

Learning strategies

- Transfer Learning





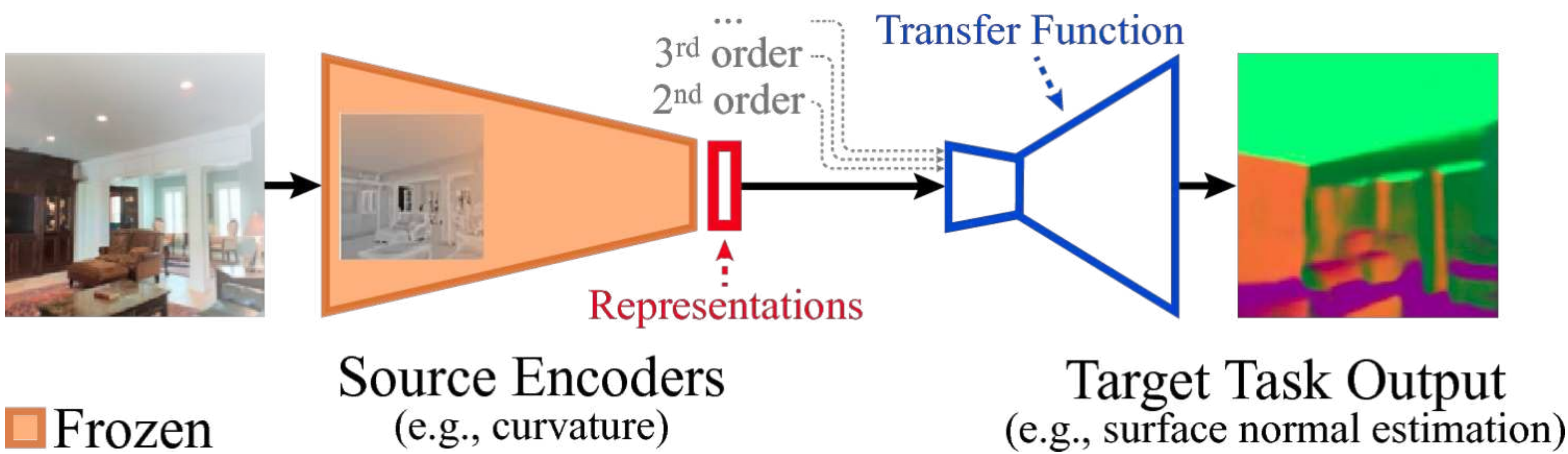
Zamir, Taskonomy: Disentangling Task Transfer Learning CVPR 2018

Learning strategies

- Transfer learning

  Consider a set of 26 typical computer vision tasks.
  Use model trained for task A to solve task B.

  1. Train full model for task A
  2. For the other 25 tasks, freeze encoder and train decoder.
  3. Performance represents the degree of dependence.



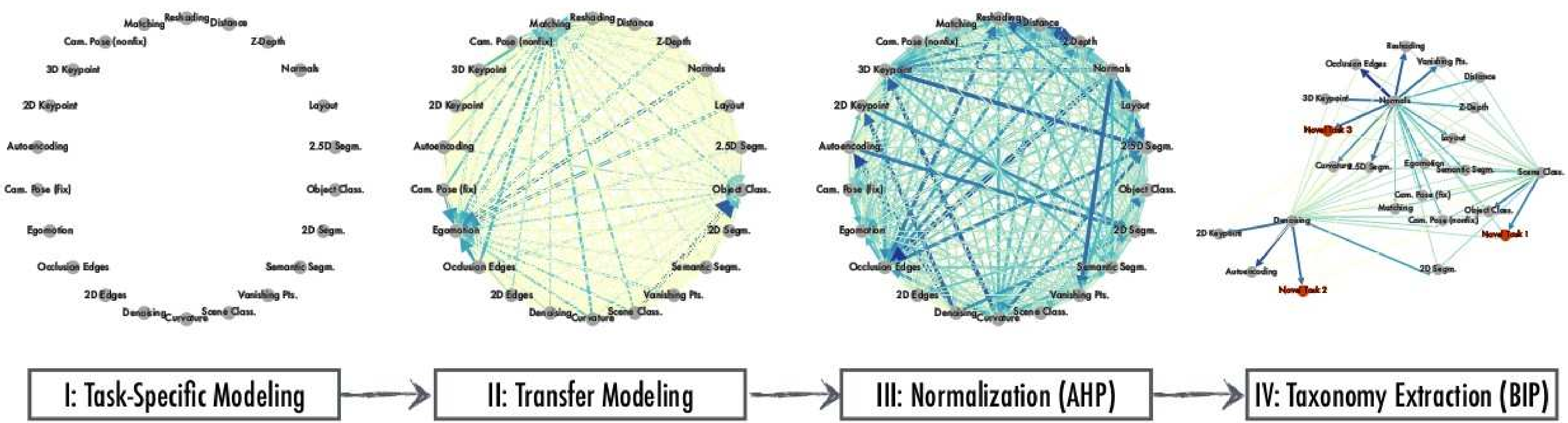
Zamir, Taskonomy: Disentangling Task Transfer Learning CVPR 2018

Learning strategies

- Transfer learning

  Bulid a graph relating tasks (nodes) with their dependencies (arrows).

  Taskonomy



A computational method for quantifying task relationships.

Exploit relationships for transfer learning.

Zamir, Taskonomy: Disentangling Task Transfer Learning CVPR 2018

Learning strategies

- Multi-task learning

  Learning strategy in which multiple learning tasks are solved at the same time.

  Builds a shared representation, so what is learned for each task can help other tasks learn better. So, potentially, it may achieve:
  - Decreased training time
  - Decreased inference time
  - More compact models
  - Increased prediction accuracy
  - Increased sample efficiency
  - Better learned representations

Standley, Which Tasks Should Be Learned Together in Multi-task Learning? ICML 2020

Learning strategies

- Multi-task learning

  However, it is not always better than single-task learning:
  negative transfer

  It depends on the relationship between the learned tasks:
  - Tasks may learn at different rates
  - One task may dominate learning
  - Gradients may interfere
  - The optimization landscape may be more difficult

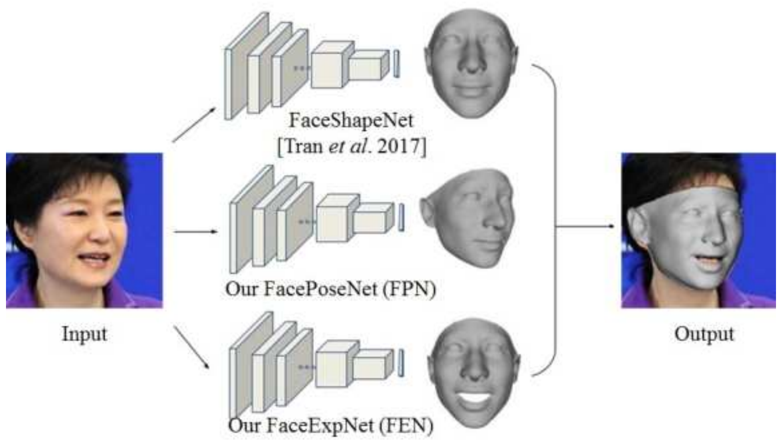
  And other features such as
  - Network size
  - Task location within the architecture

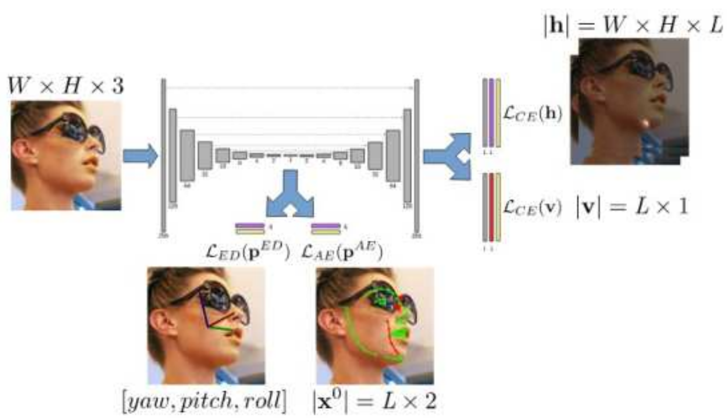
Standley, Which Tasks Should Be Learned Together in Multi-task Learning? ICML 2020

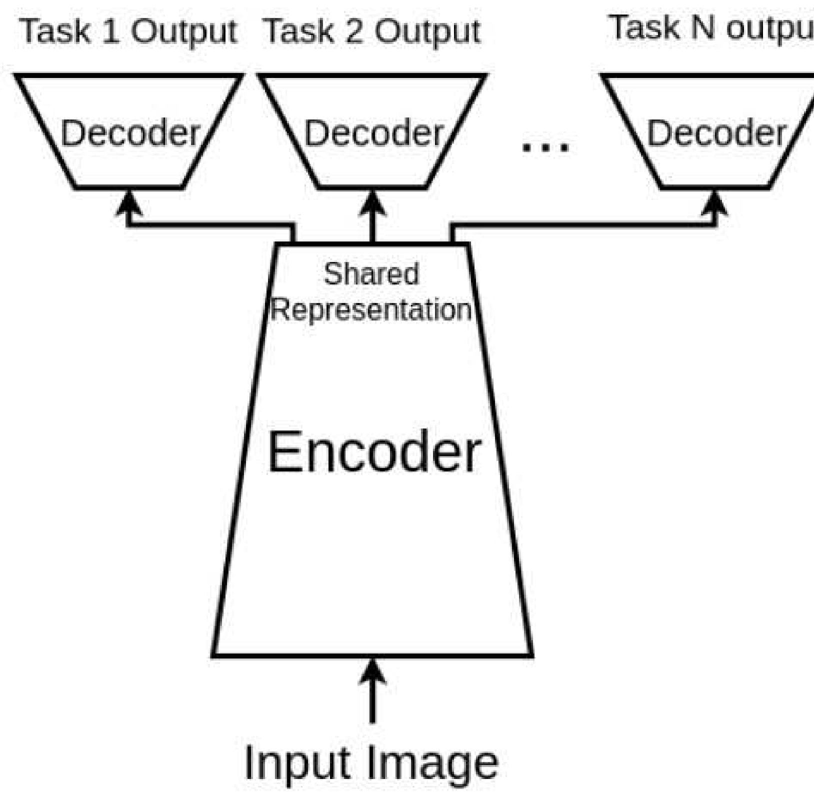
Learning strategies

- Multi-task learning

  Tasks may be allocated in different ways:





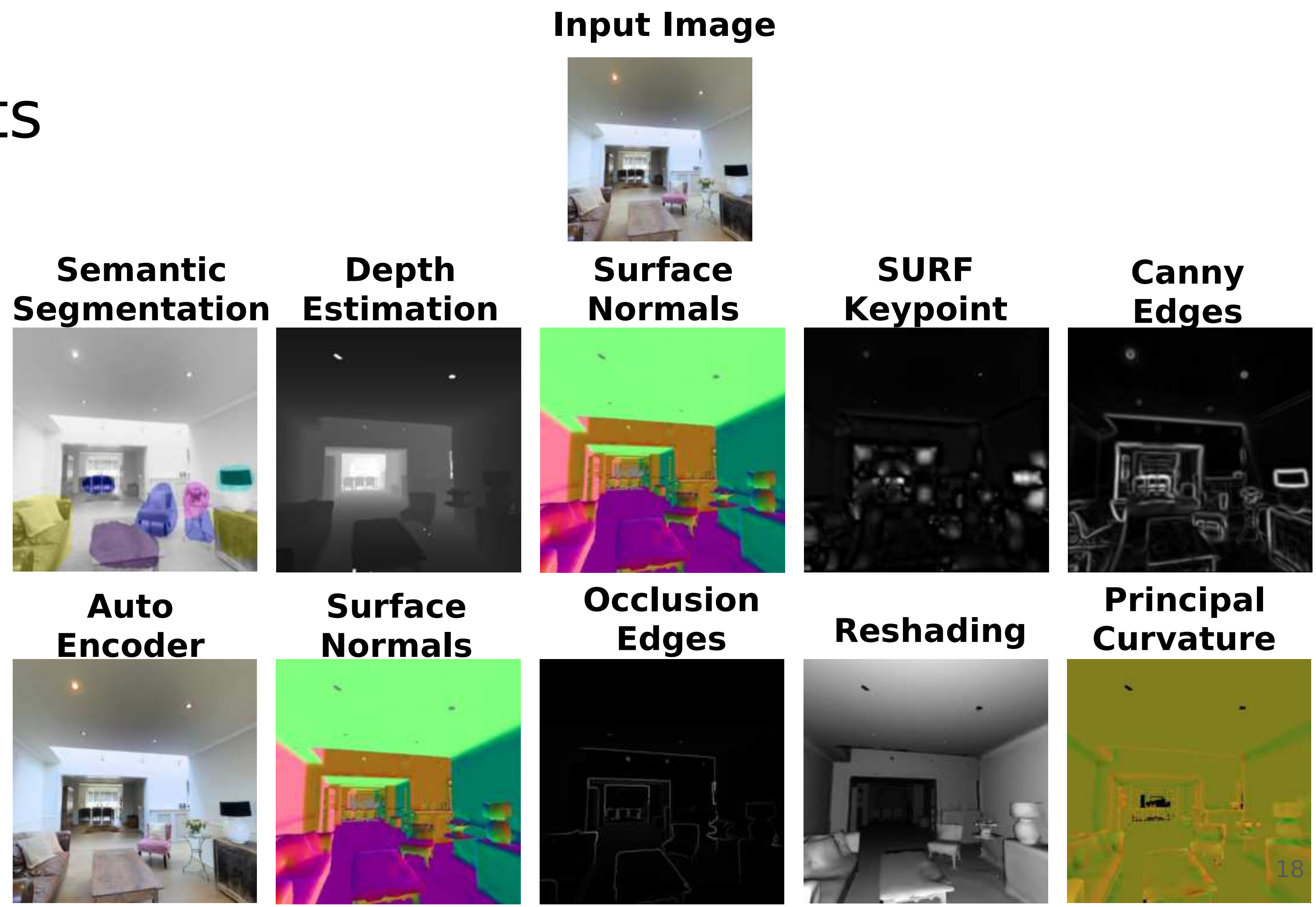


Learning strategies

- Multi-task learning. Practical applications

Indoor image analysis

Task Sets

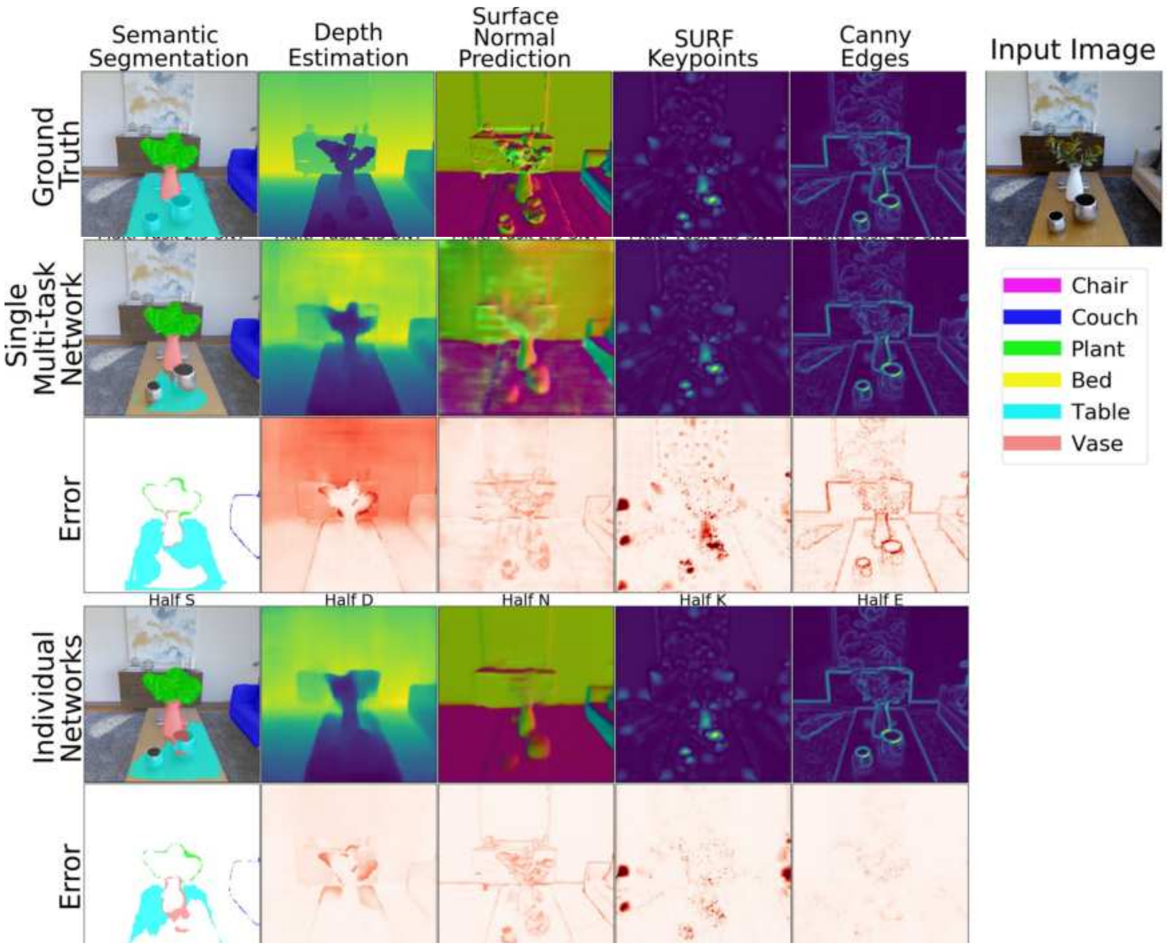


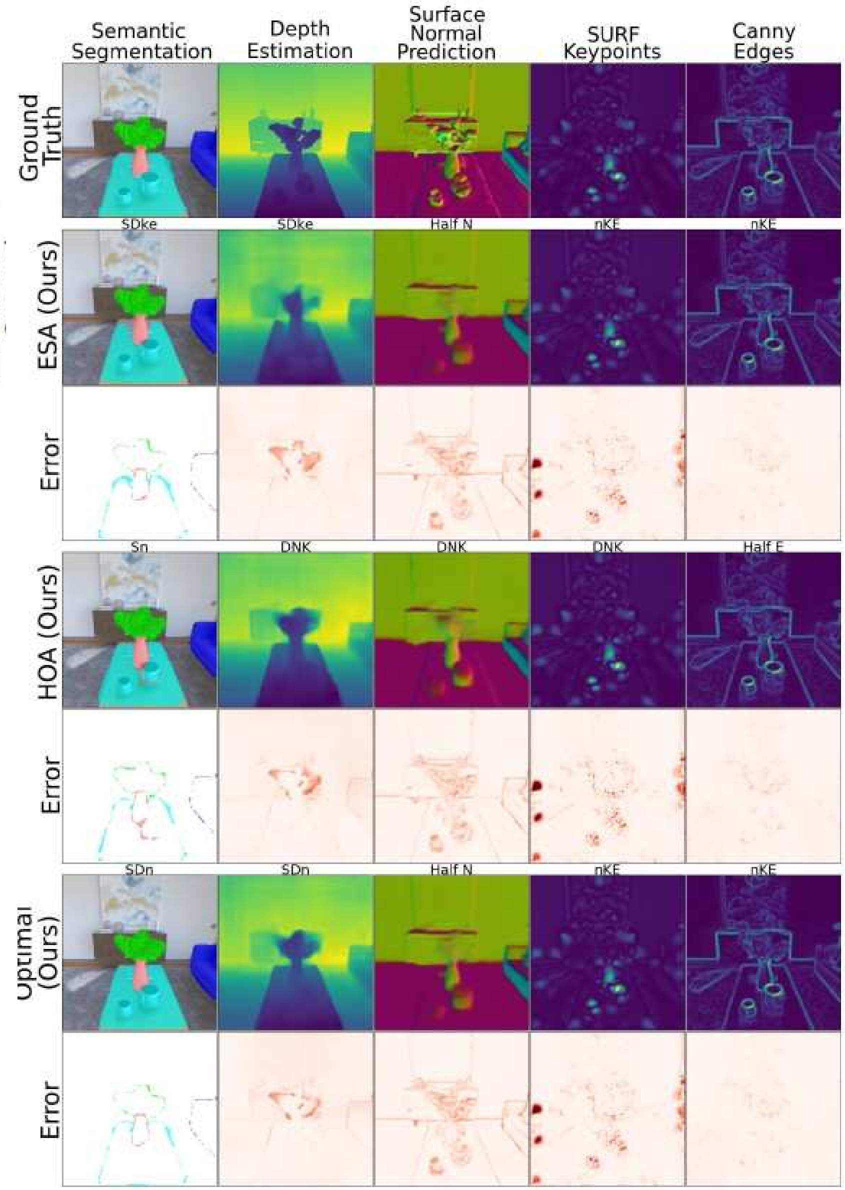
Standley, Which Tasks Should Be Learned Together in Multi-task Learning? ICML 2020

Learning strategies

- Multi-task learning. Practical applications

  Qualitative results





Standley, Which Tasks Should Be Learned Together in Multi-task Learning? ICML 2020

Learning strategies

- Multi-task learning. Practical applications

  Conclusions

  Many common assumptions do not seem to be true
  - Similar tasks don't necessarily work better together
  - No a priori way to tell which tasks will work well together
  - MTL doesn't necessarily work better when you have less data
  - Task relationships are not the same between settings. They are sensitive to:
    - ➔ Dataset size
    - ➔ Network capacity
    - ➔ Location of tasks in the architecture

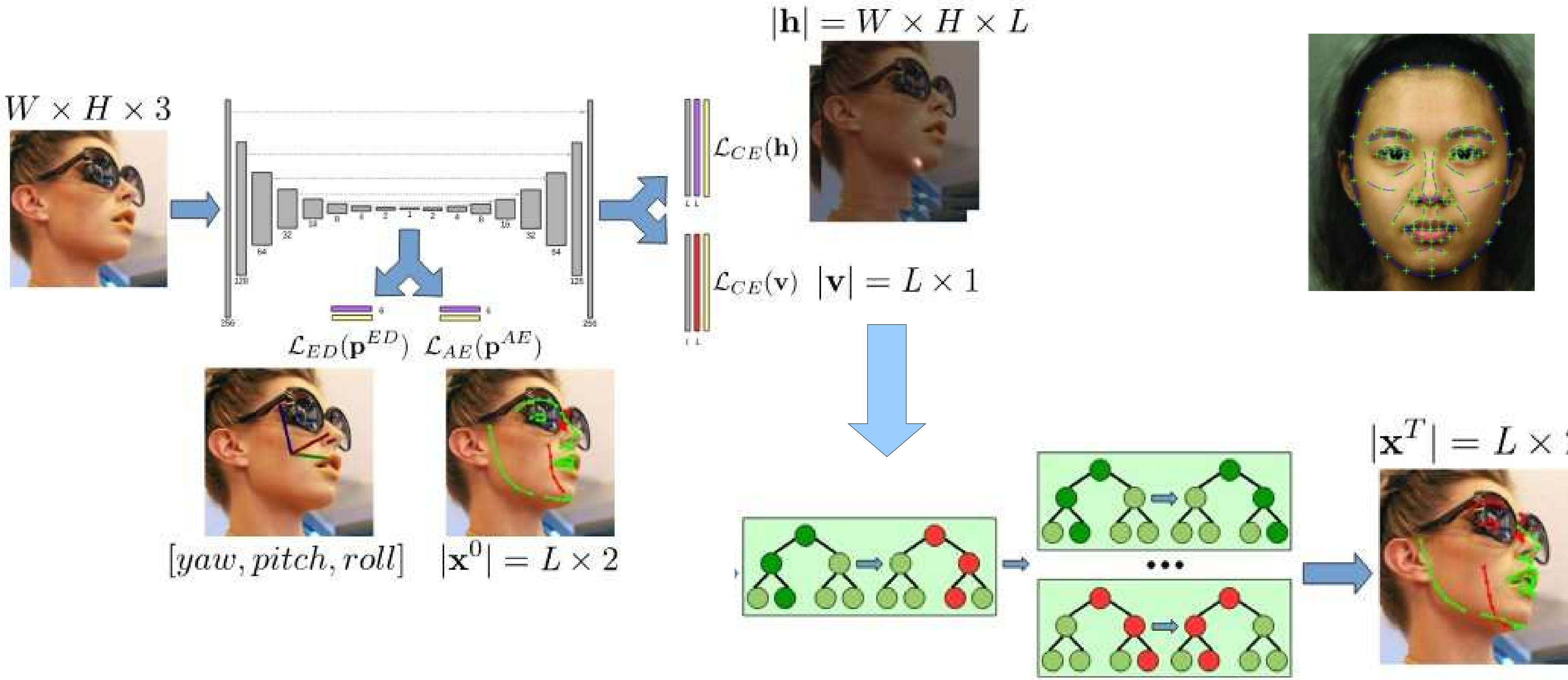
    … and probably other variables

Standley, Which Tasks Should Be Learned Together in Multi-task Learning? ICML 2020

Learning strategies

- Multi-task learning. Practical applications

Head pose, landmarks and visibility estimation

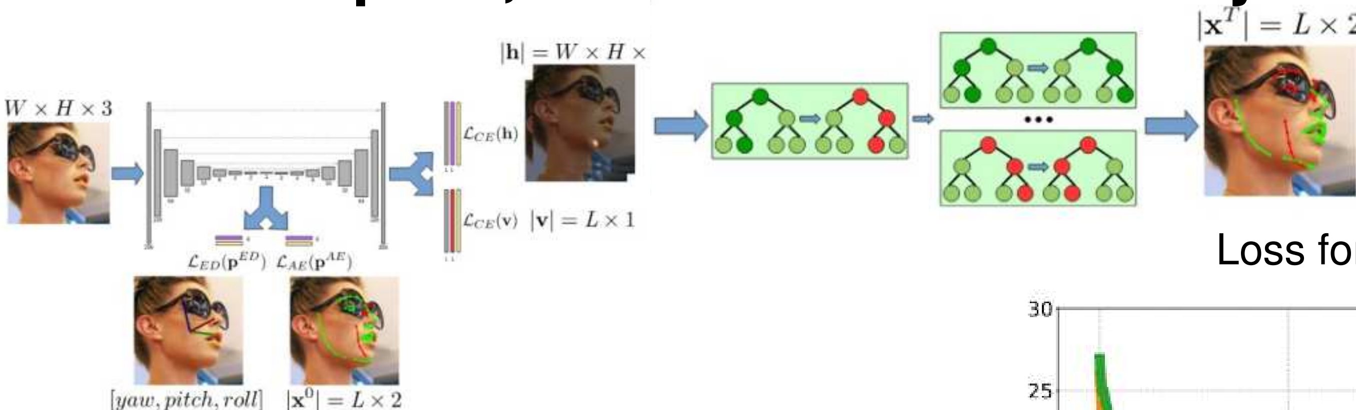


Valle, Multi-task head pose estimation in the wild, PAMI 2021

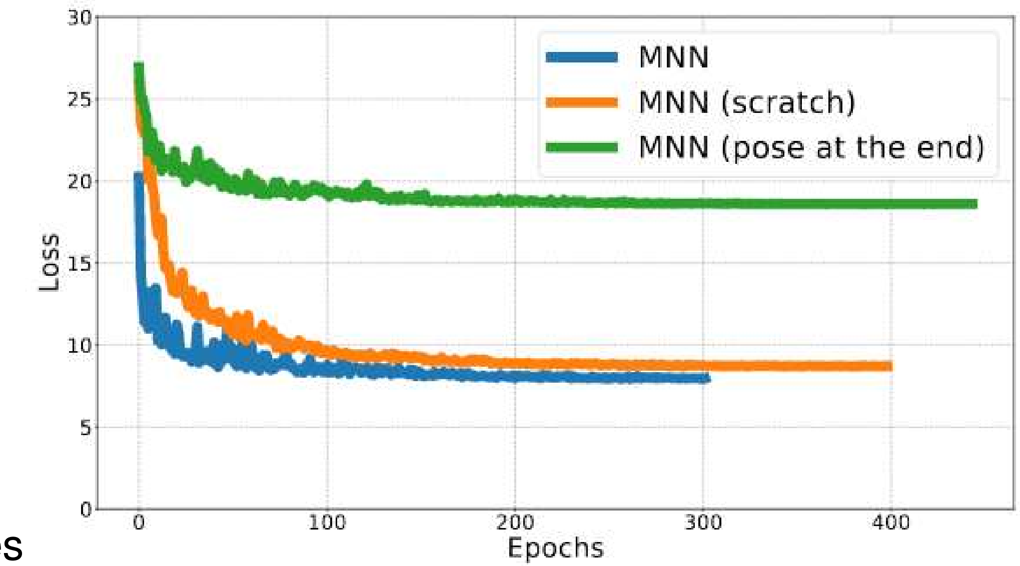
Learning strategies

- Multi-task learning. Practical applications

Head pose, landmarks and visibility estimation



Loss for different task locations



NME for different training strategies

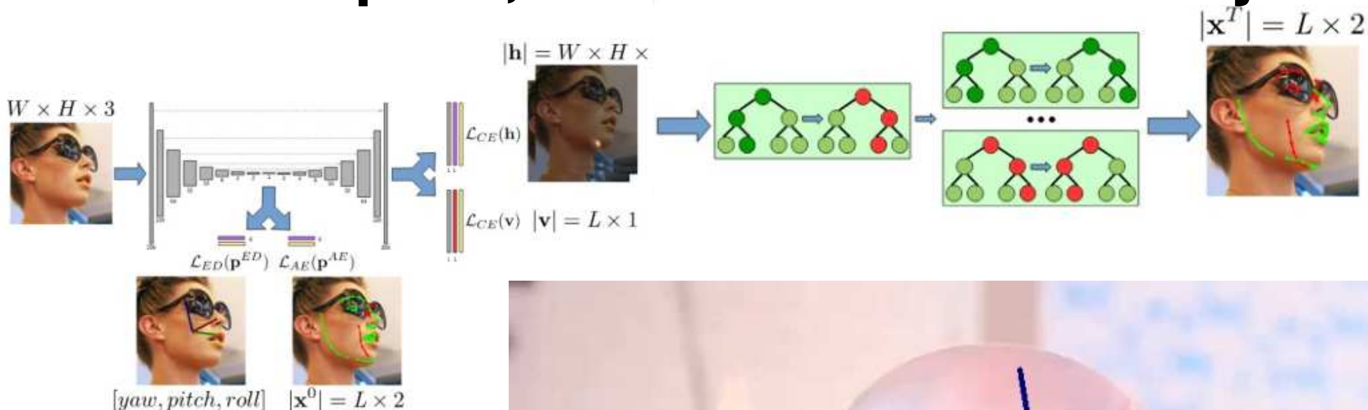
Method		300W pub	300W priv	COFW	AFLW	WFLW	Avg
Single task	Pose	1.91	2.22	2.67	3.43	2.46	2.54
Multi-task	Sym	1.76	1.97	2.57	3.35	2.10	2.35
	Pre+Sym	1.59	1.96	2.36	3.22	2.08	2.24
	Pre+Pose	1.56	1.96	2.34	3.23	2.11	2.24

Valle, Multi-task head pose estimation in the wild, PAMI 2021

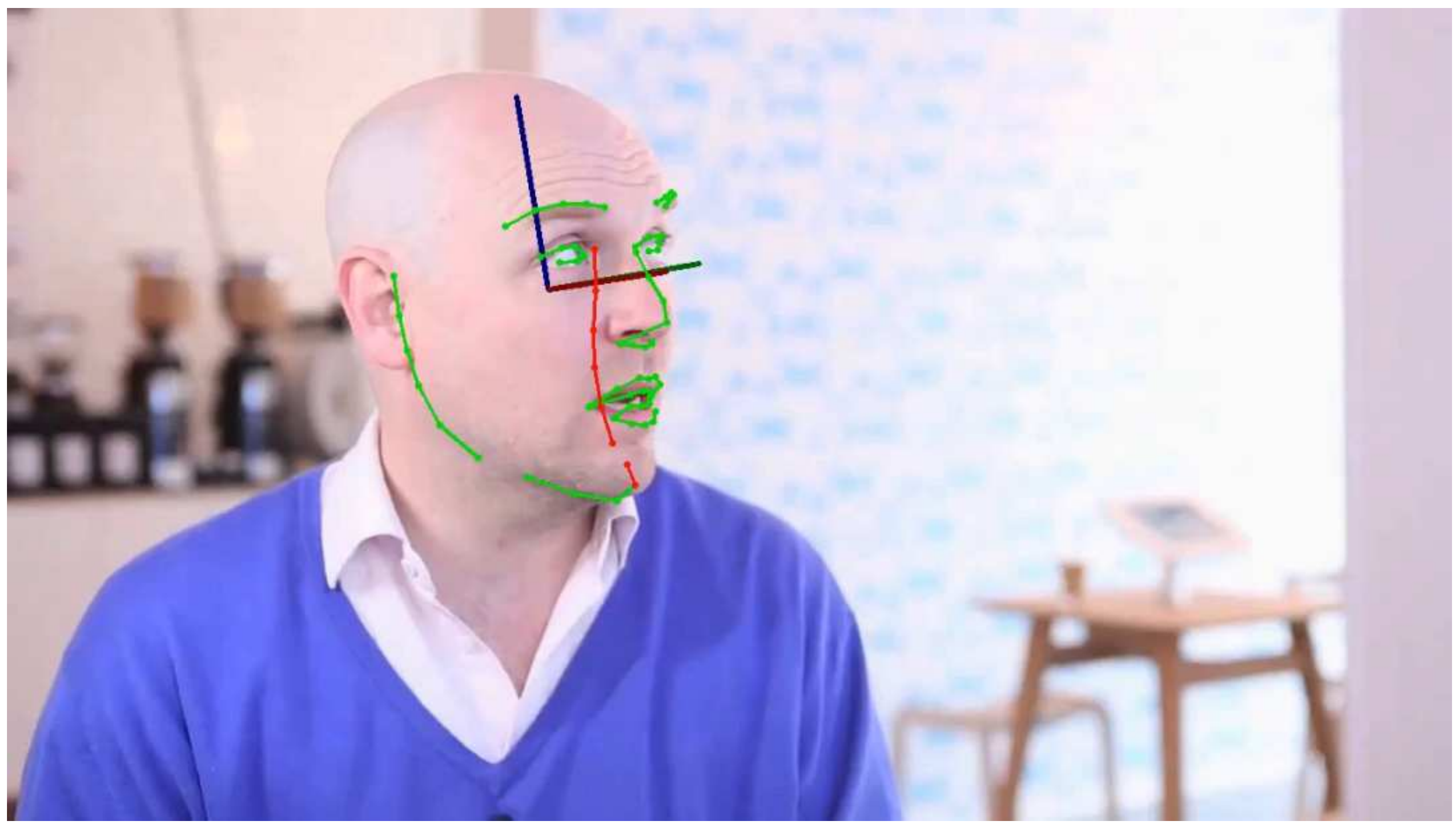
Learning strategies

- Multi-task learning. Practical applications

Head pose, landmarks and visibility estimation



Qualitative results

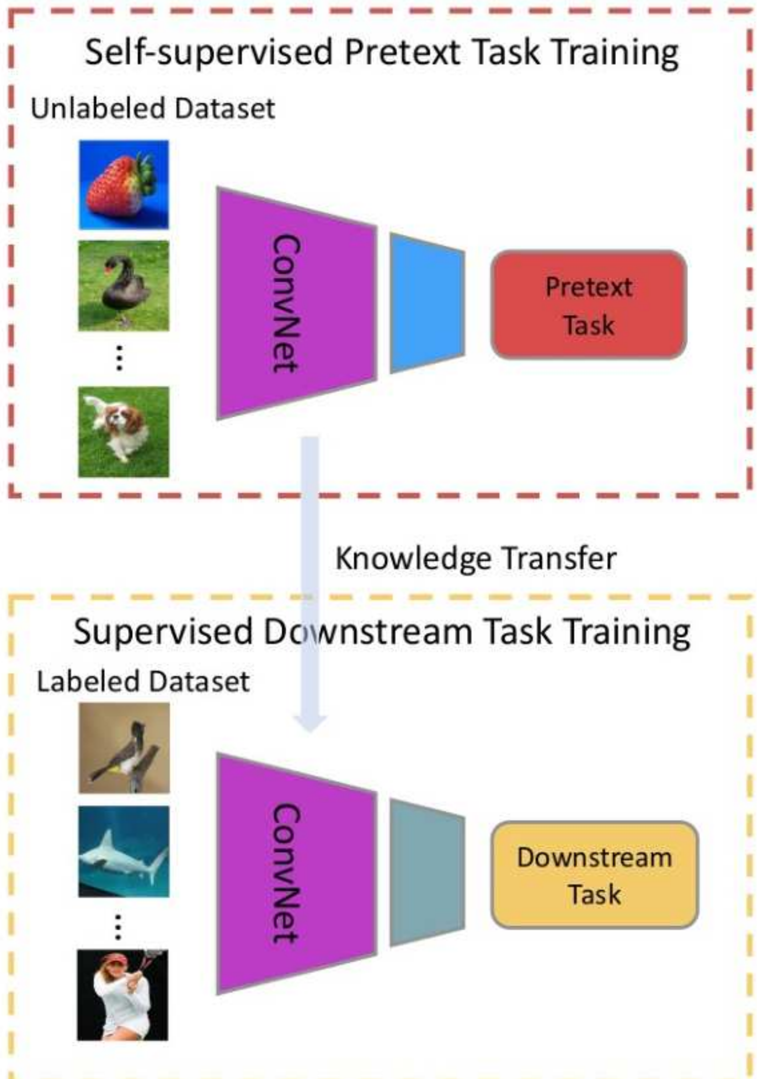


Valle, Multi-task head pose estimation in the wild, PAMI 2021

Learning strategies

- Self-supervised learning

The model is trained with automatically generated tasks labels.



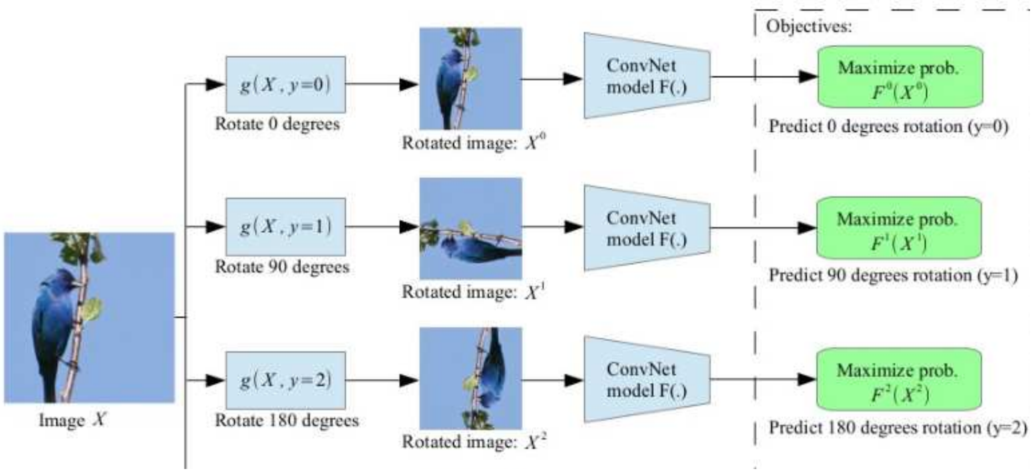
- tasks difficult to solve so they produce a good representation
- labels to formulate the objective function obtained automatically

Types of tasks:
- Image
- Video
- Multi-modal
- Synthetically generated

Jing, Self-supervised Visual Feature Learning with Deep Neural Networks, PAMI 2021

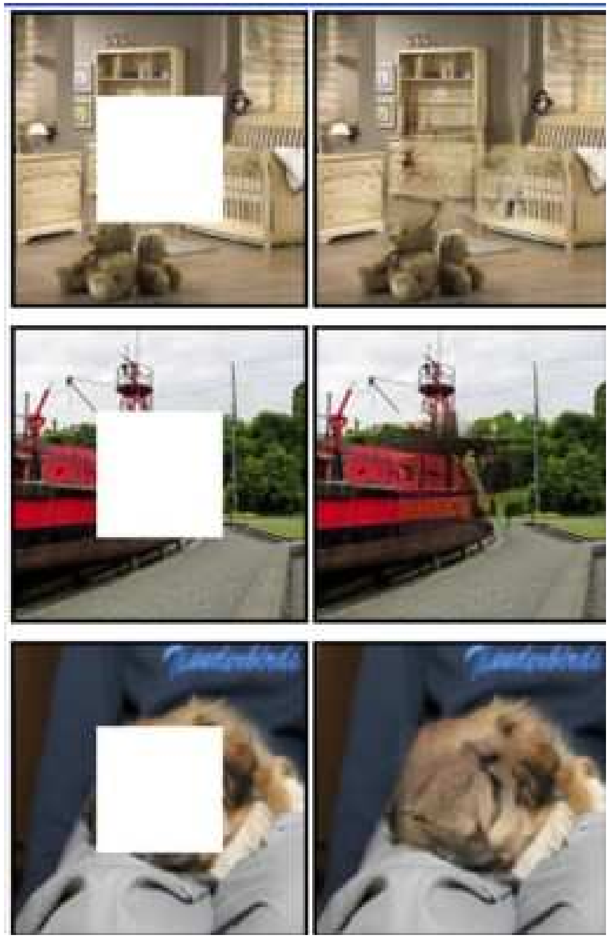
Learning strategies

- Self-supervised learning. Image-based tasks



Gidaris ICLR18 - Image rotation prediction

Zhang ECCV16 - Colorization

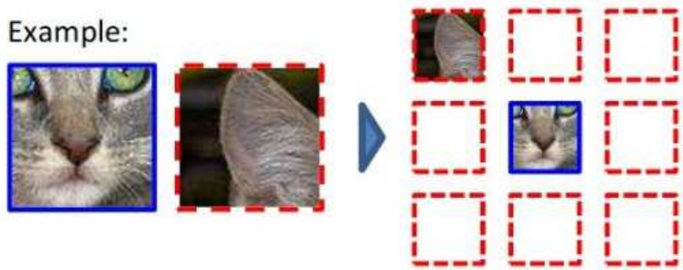


Pathak CVPR16 - Image inpainting

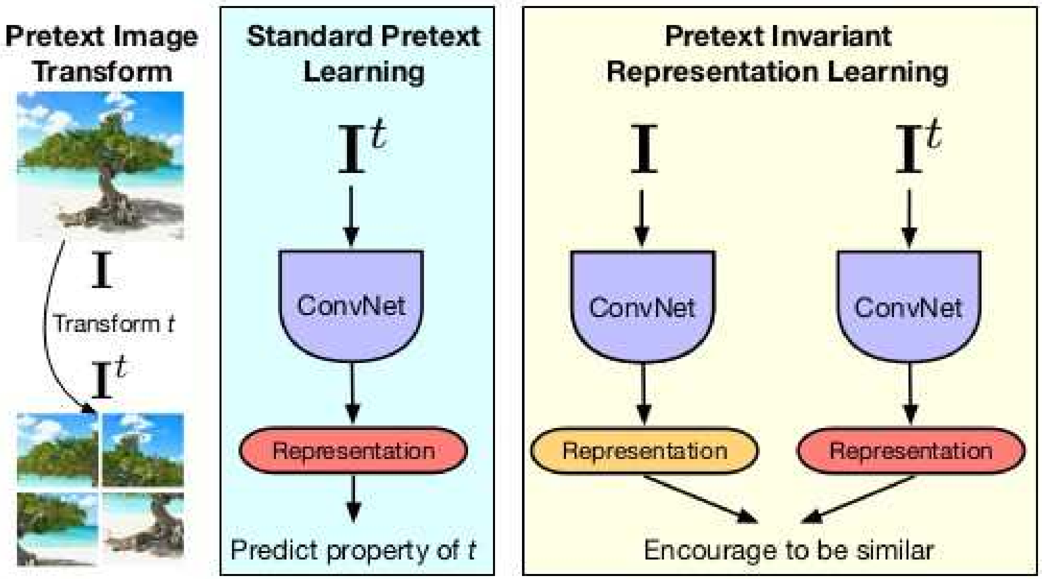
Learning strategies

- Self-supervised learning. Image-based tasks

Context related tasks



Doersch ICCV15 Patch location prediction



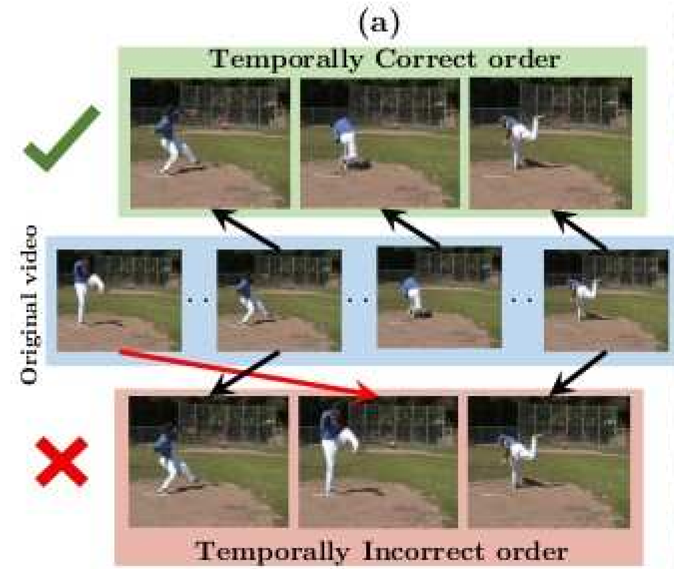
Misra CVPR20 – Pretext-invariant Representations

Noroozi ECCV16 – Solving jigsaw puzzle

Learning strategies

- Self-supervised learning. Video-based tasks

Aggrawal ICCV15 Learning to See by Moving



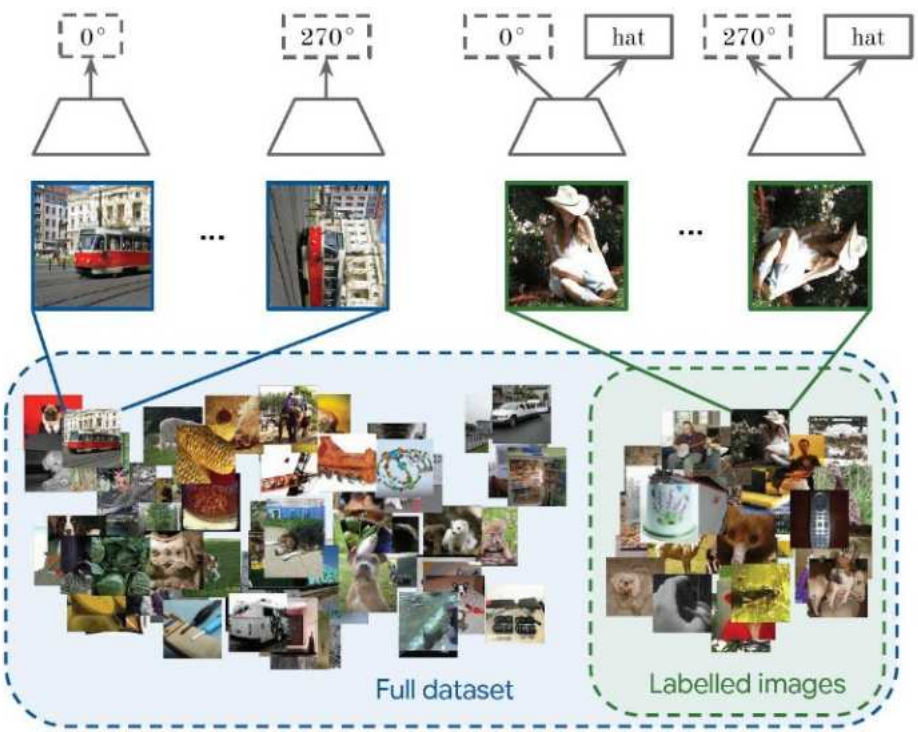
Misra ECCV16-Temporal Order Verification

Learning strategies

- Semi-supervised and self-supervised learning

  Use a small amount of labeled data with a large amount of unlabeled data.

  Semi-supervised + self-supervised



Self Supervised:

Apply any self-supervision loss on data without labels

$\mathcal{L}_u = \mathcal{L}_{rot} \,/\, \mathcal{L}_{exemplar}$

Semi Supervised:

Apply regular supervised loss on data with labels

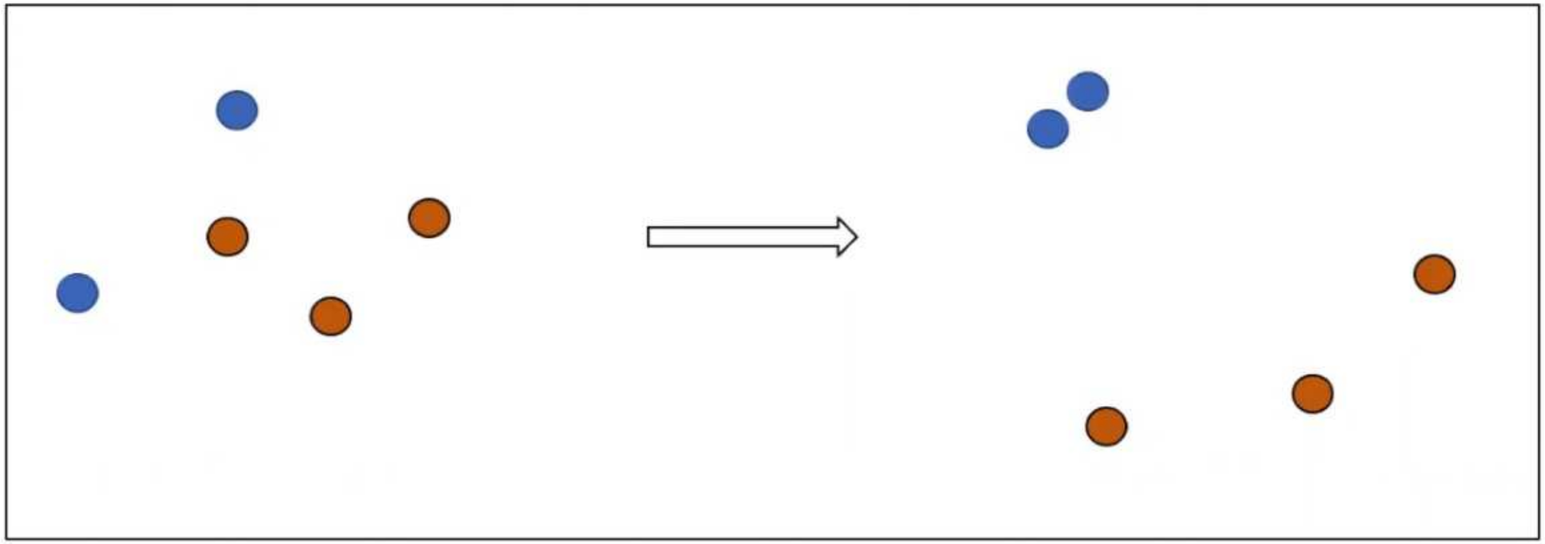
$\mathcal{L} = w_{sup}\, \mathcal{L}_{sup} + w_{rot}\, \mathcal{L}_{rot}$

Zhai, S4L: Self-Supervised Semi-Supervised Learning ICCV 2019

Learning strategies

- Self-supervised learning. **Contrastive learning**
  The contrastive task

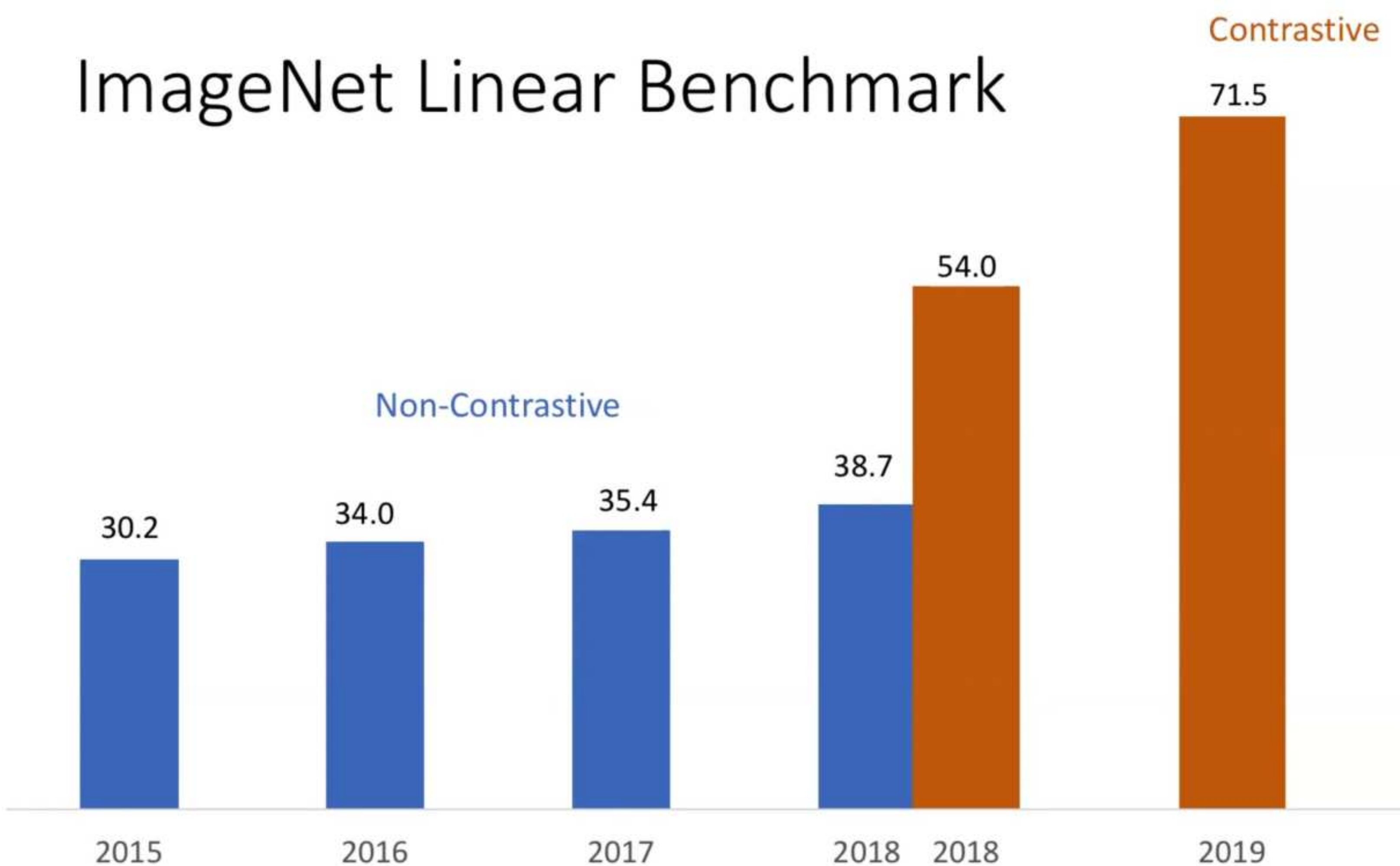
• Pulling together similar pairs, Push away dissimilar pairs



Tian - Contrastive Learning: A General Self-supervised Learning Approach, 2020

Learning strategies

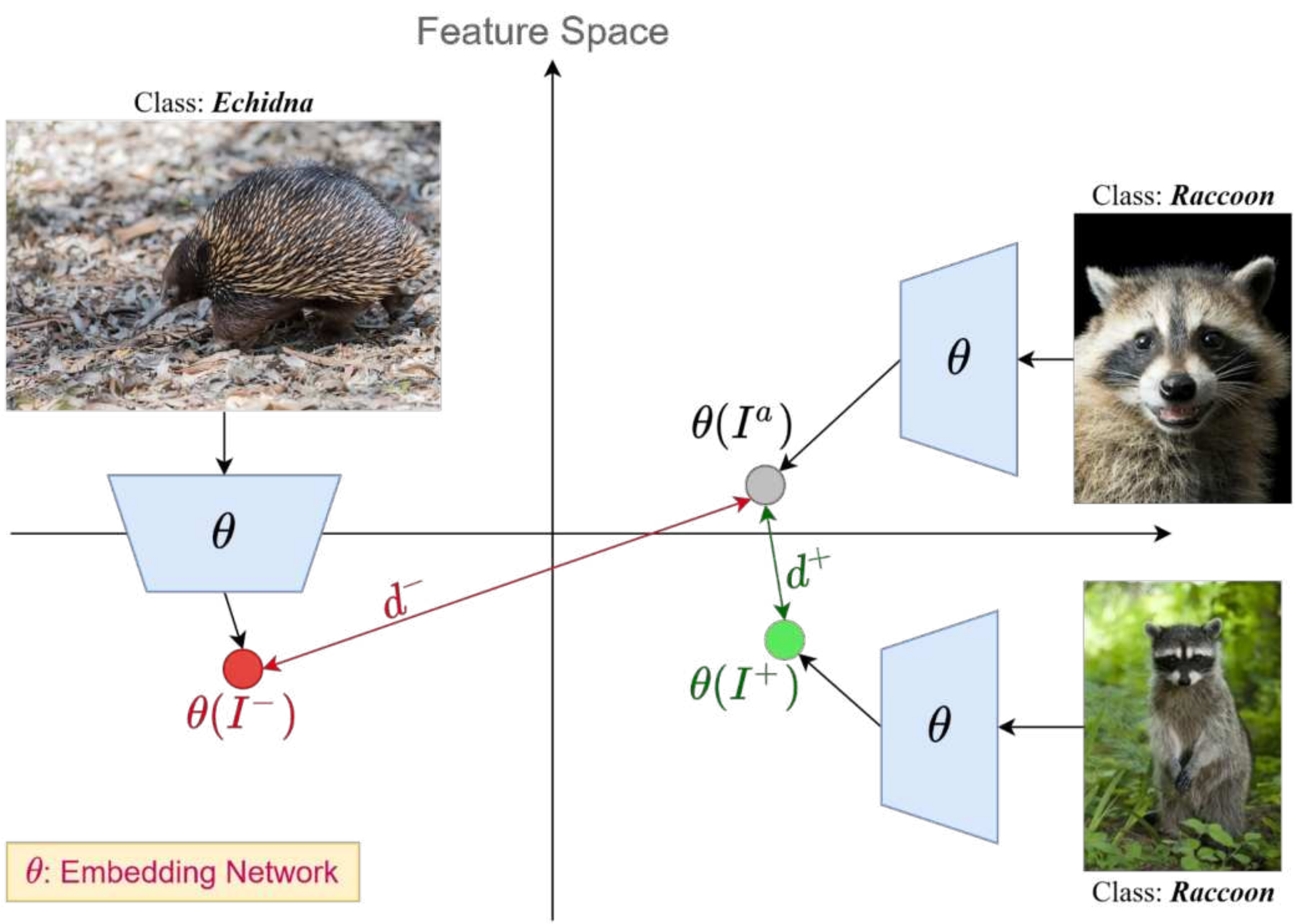
- Contrastive learning
  ImageNet unsupervised classification performance



Tian - Contrastive Learning: A General Self-supervised Learning Approach, 2020

Learning strategies

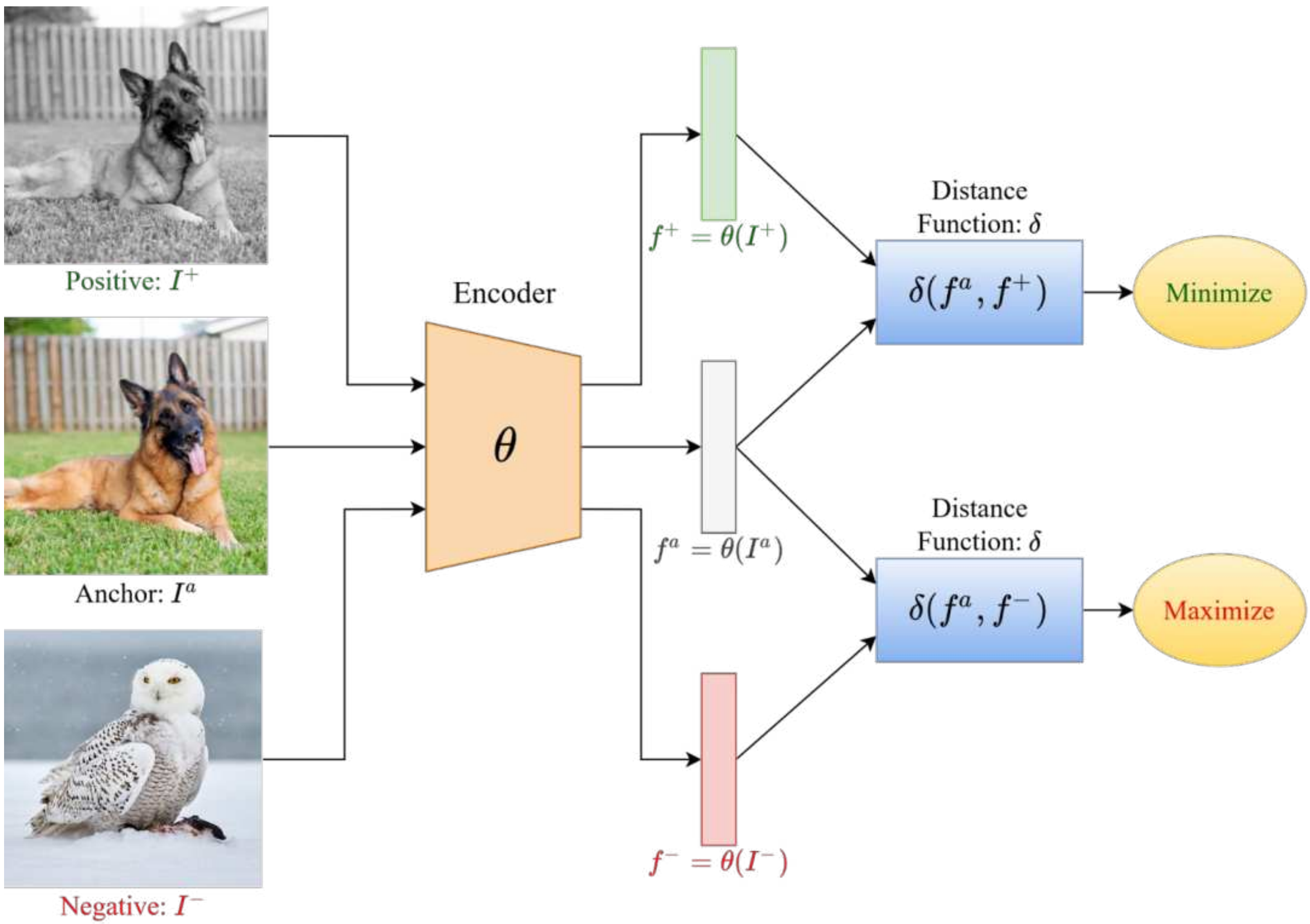
- Self-supervised learning. **Contrastive learning**
  The contrastive task



https://www.v7labs.com/blog/contrastive-learning-guide

Learning strategies

- Self-supervised learning. **Contrastive learning**
  The self-supervised contrastive task

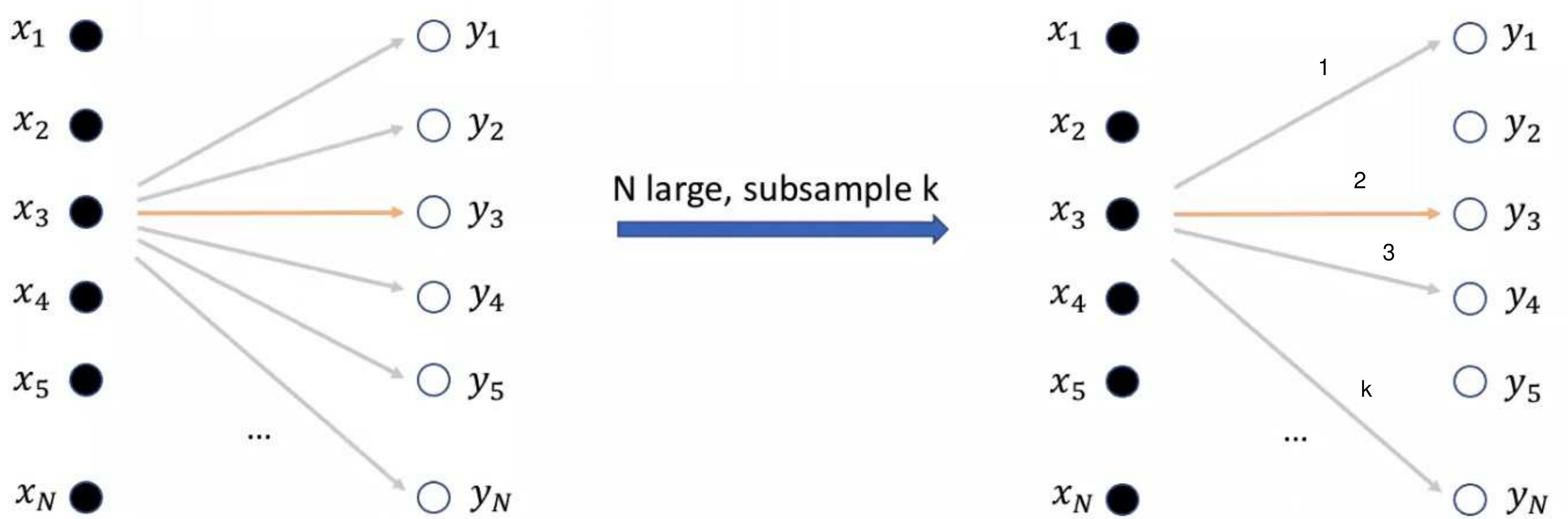


https://www.v7labs.com/blog/contrastive-learning-guide

Learning strategies

- Contrastive learning

• A set pf paired samples $\{x_i, y_i\}_{i=1}^{N}$



$$\ell = -\log \frac{\exp(\mathrm{sim}(x_+, y_+)/\tau)}{\exp(\mathrm{sim}(x_+, y_+)/\tau) + \sum_{i=1}^{k} \exp(\mathrm{sim}(x_+, y_-^i)/\tau)}$$

N-pair contrastive Loss

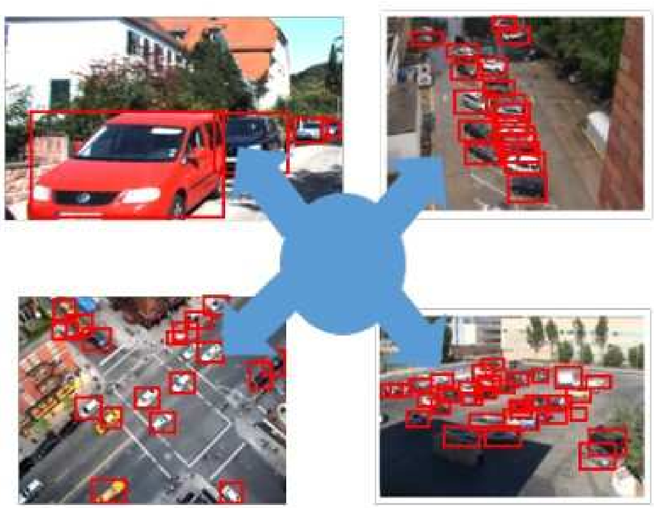
Tian - Contrastive Learning: A General Self-supervised Learning Approach, 2020

Learning strategies

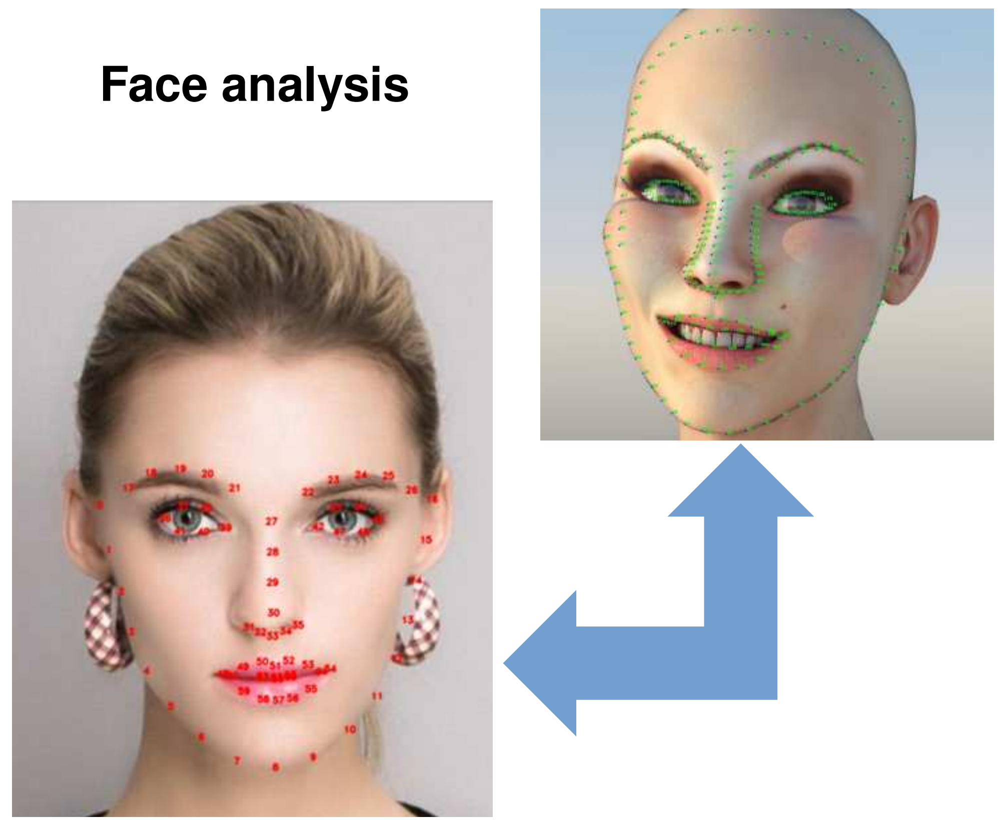
- Domain adaptation

  Leverages labeled data in source domains to train a model with labeled or unlabeled data in a different target domain.

Object detection



Face analysis



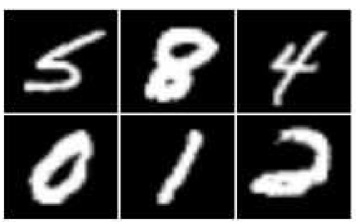
G. Csurka, Domain Adaptation for Visual Applications: A Comprehensive Survey 2017

Learning strategies

- Domain adaptation

Leverages labeled data in source domains to train a model with labeled or unlabeled data in a different target domain.

Source domain
labeled data



MNIST

+

Target domain
unlabeled data

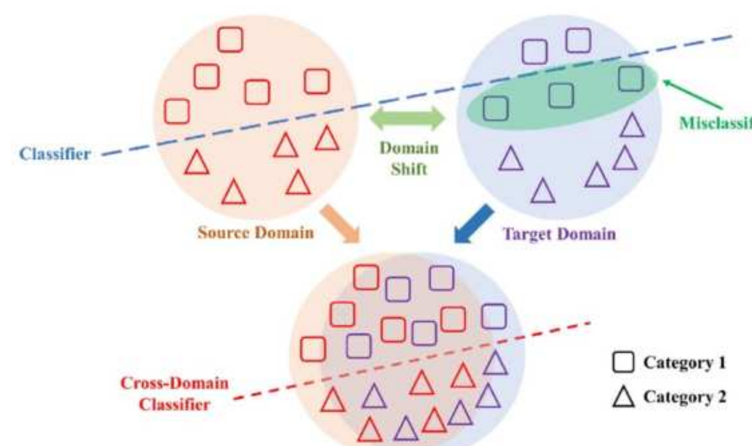
SVHN

Training data

Unsupervised
Domain adaptation

SVHN

Domain shift produces a mismatch between the distributions of train and test data, degrading the performance.

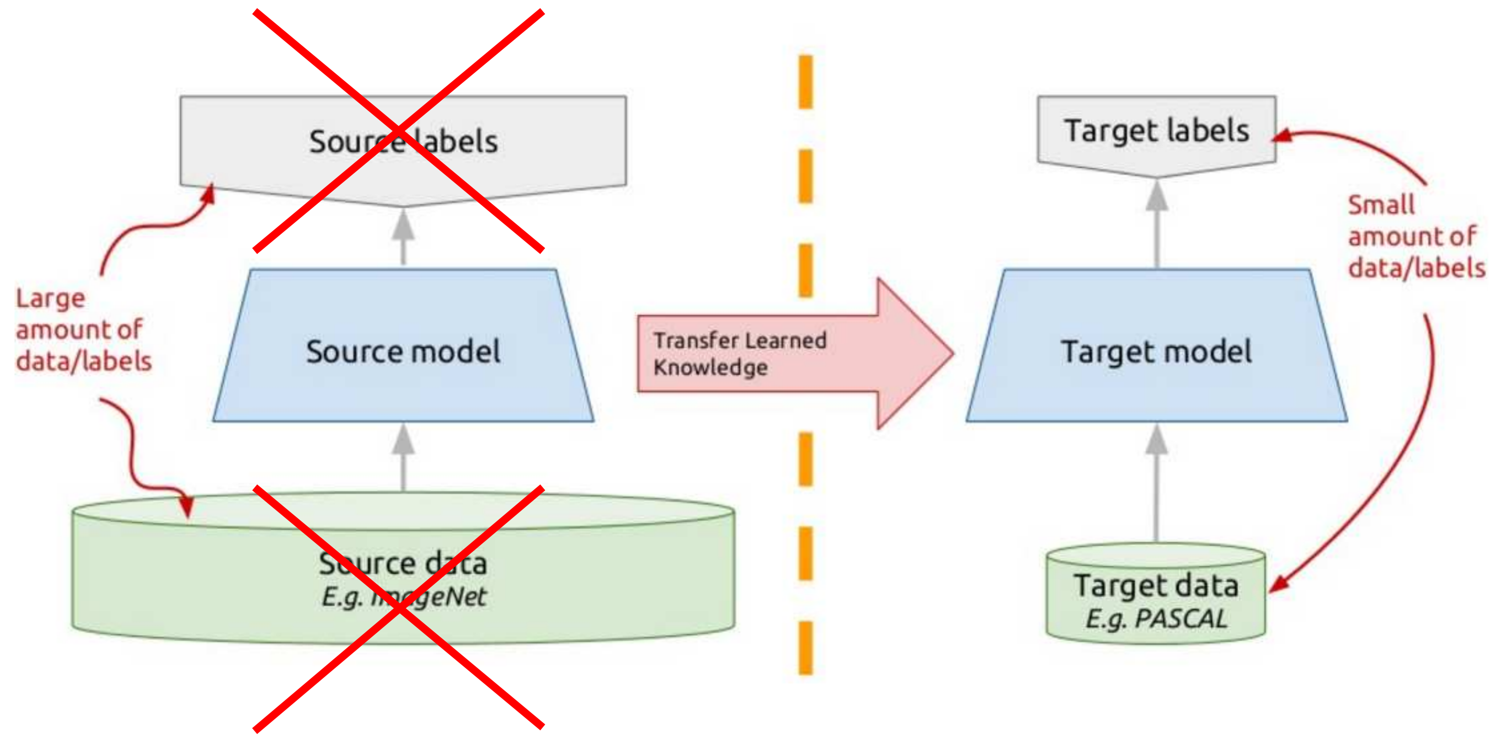


Learning strategies

- Types of domain adaptation (DA) techniques

  Depending on the *availability of target data*, DA techniques may be organized into:

  - Available and annotated.
    Transfer learning.
    The model is usually pretrained



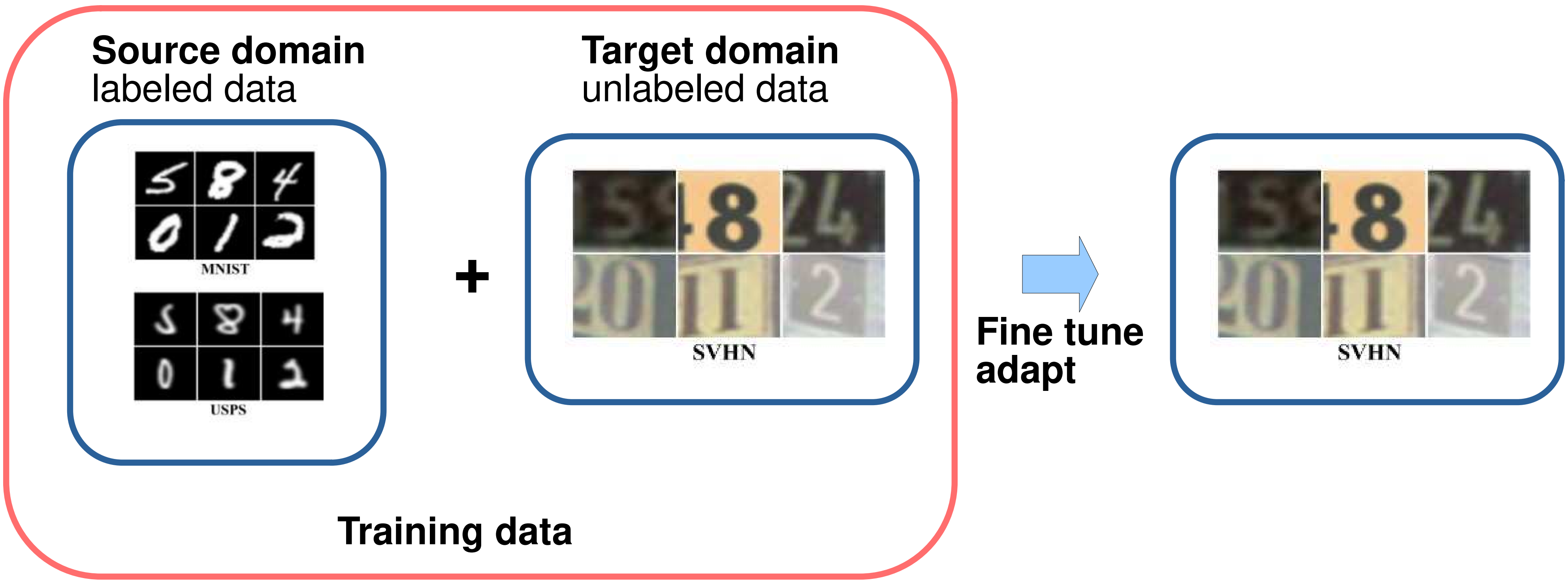
Learning strategies

- Types of domain adaptation (DA) techniques

  Depending on the *availability of target data*, DA techniques may be organized into:

  - Available and annotated.
  - Available unannotated.

    Unsupervised domain adaptation



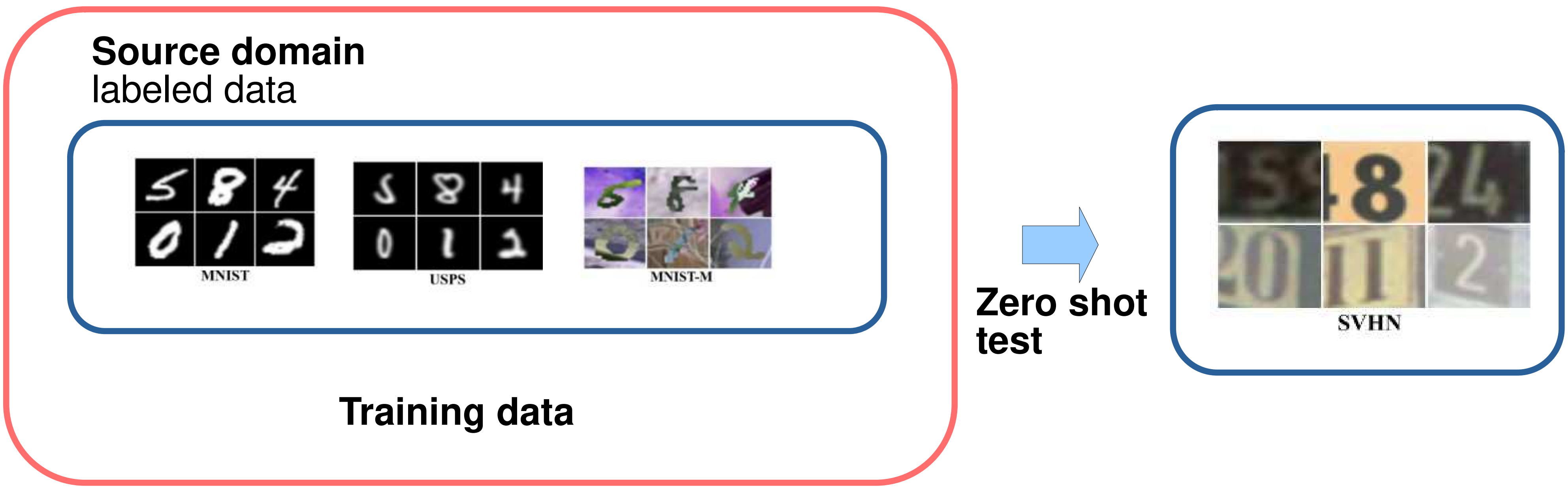
Learning strategies

- Types of domain adaptation (DA) techniques

  Depending on the *availability of target data*, DA techniques may be organized into:

  - Available and annotated.
  - Available unannotated
  - Unavailable

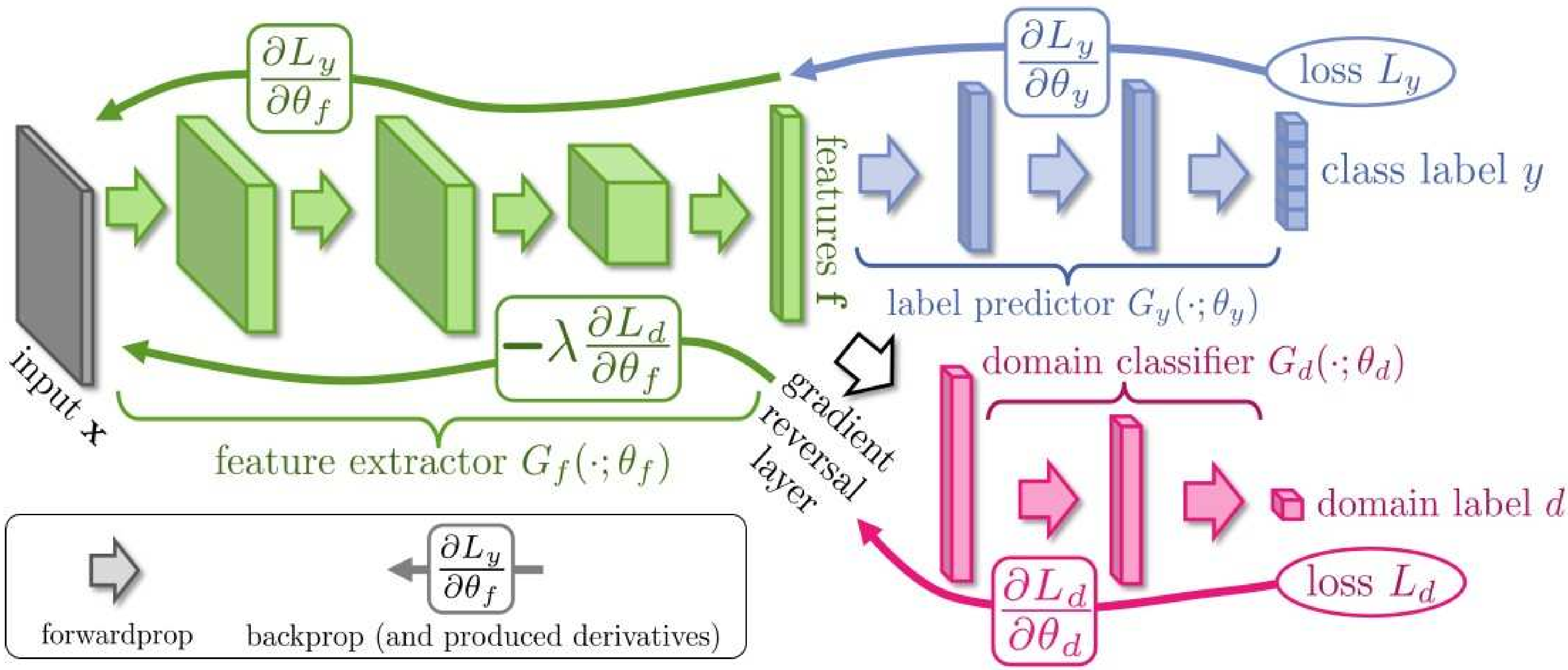
    Domain generalization



Learning strategies

- Un supervised domain adaptation

Adversarial domain adaptation



Ganin, Domain-Adversarial Training of Neural Networks, JMLR 2017

Learning strategies

- Un supervised domain adaptation

Adversarial domain adaptation

Method (Source / Target)	MNIST / MNIST-M	Syn Numbers / SVHN	SVHN / MNIST	Syn Signs / GTSRB
Source only	.5225	.8674	.5490	.7900
SA (Fernando et al., 2013)	.5690 (4.1%)	.8644 (−5.5%)	.5932 (9.9%)	.8165 (12.7%)
DANN	**.7666** (52.9%)	**.9109** (79.7%)	**.7385** (42.6%)	**.8865** (46.4%)
Train on target	.9596	.9220	.9942	.9980

Ganin, Domain-Adversarial Training of Neural Networks, JMLR 2017

Representation Learning

- Conclusion.

- **Representation learning** is a central concept in the development of machine learning for AI.
- Representations may be reused. Having a good representation is an excellent starting point for a solution to any problem.
- Annotated data is scarce, so
  - Transfer learning
  - Multi-task learning
  - Unsupervised learning
  - Domain Adaptation
  - Domain Generalization

  are tools to address the representation learning problem.