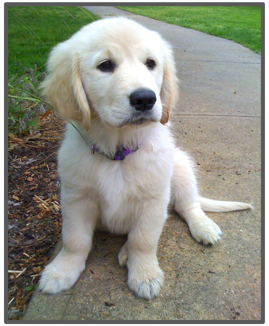
Self Supervision Techniques in CNNs

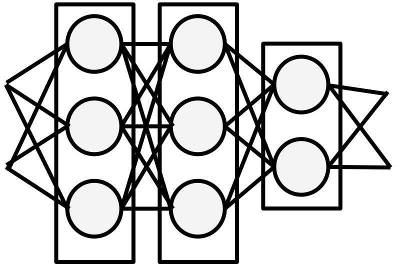
Varsha S
193079005

Outline

1. Motivation
2. Self supervision
3. Pretext Task
    a. Inpainting
    b. Jigsaw Puzzles

Motivation



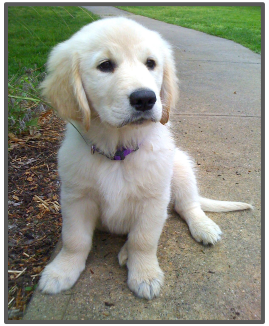


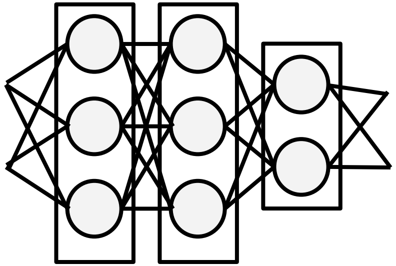
Golden Retriever

Deep learning + ImageNet

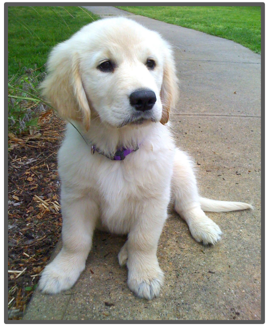
* *Image source:-Google images*

Motivation





Golden Retriever



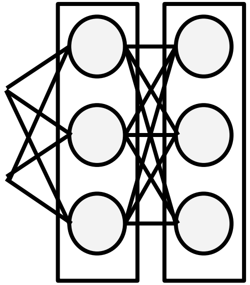
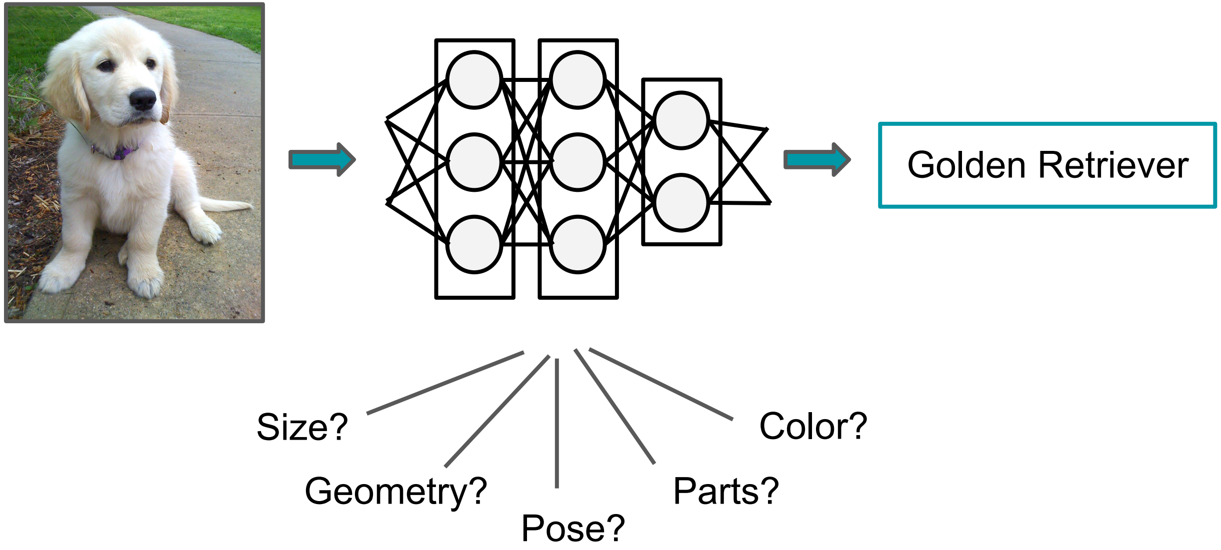


Image Segmentation
Detection
Depth Estimation
...

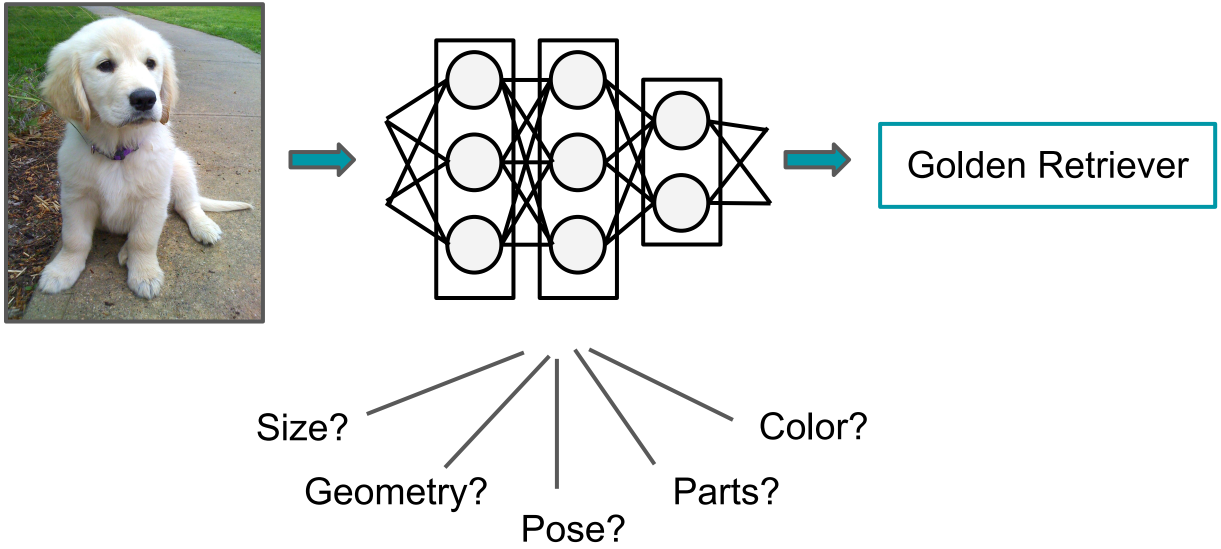
** Image source:-Google images*

Motivation



* *Image source:-Google images*

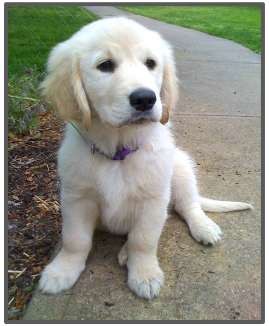
Motivation

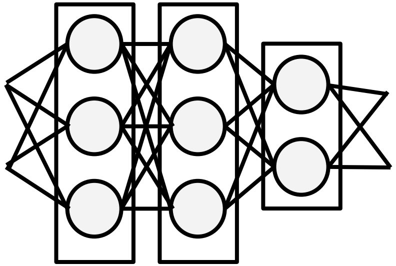


Can the task be something else ?

** Image source:-Google images*

Motivation





Golden Retriever

Size?

Geometry?

Pose?

Parts?

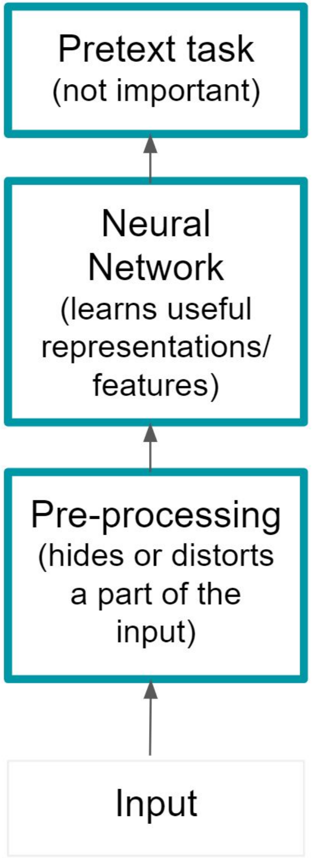
Color?

Are the labels necessary?

* *Image source:-Google images*

Self Supervision

- Data provides supervision
- Goal - Learn good representations
- Task - Design pretext



* Source - https://annotation-efficient-learning.github.io/slides/relja_cvpr2020_self_supervised.pdf

PRETEXT TASK

A. INPAINTING

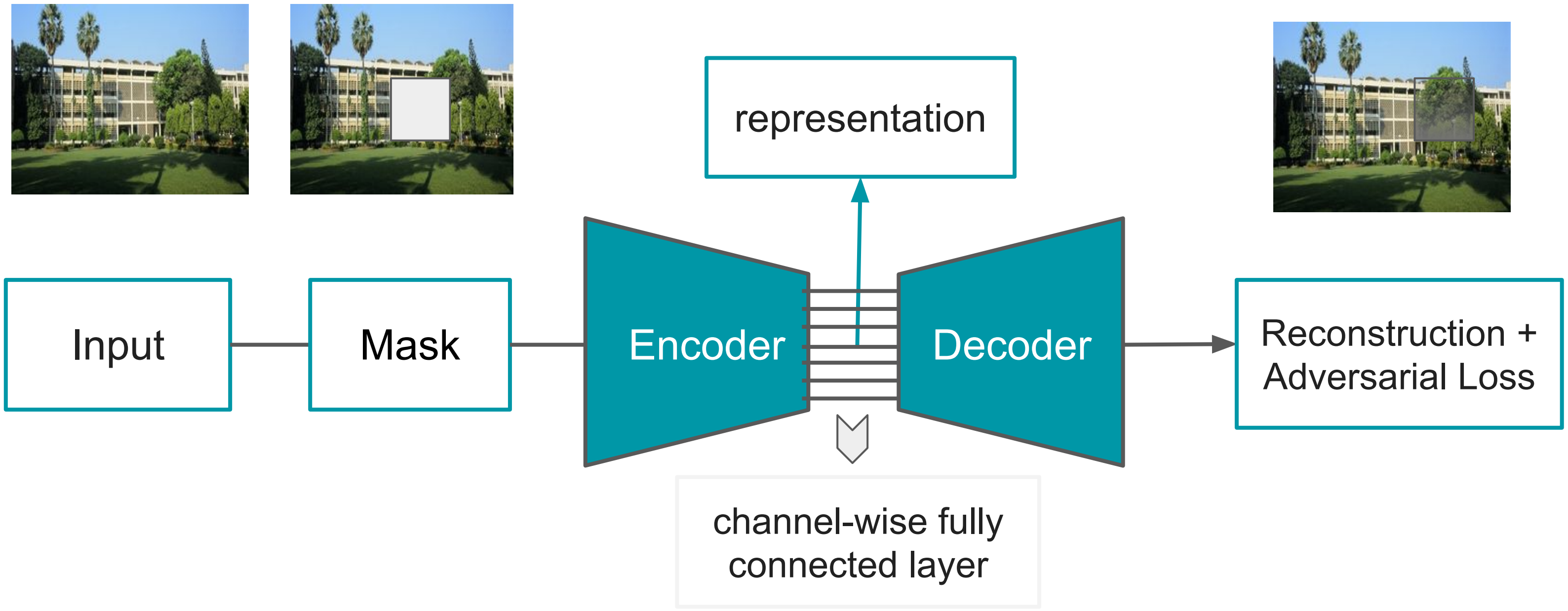
Inpainting - Context encoders

* *Image source:-Google images*

Inpainting - Context encoders

* *Image source:-Google images*

Inpainting - Context encoders



** Image source:-Google images*

Inpainting - Loss function

- $\mathcal{L} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{adv}\mathcal{L}_{adv}.$

- $\mathcal{L}_{rec}(x) = \|\hat{M} \odot (x - F((1 - \hat{M}) \odot x))\|_2^2,$

- $$\mathcal{L}_{adv} = \max_{D} \quad \mathbb{E}_{x \in \mathcal{X}}[\log(D(x)) + \log(1 - D(F((1 - \hat{M}) \odot x)))]$$

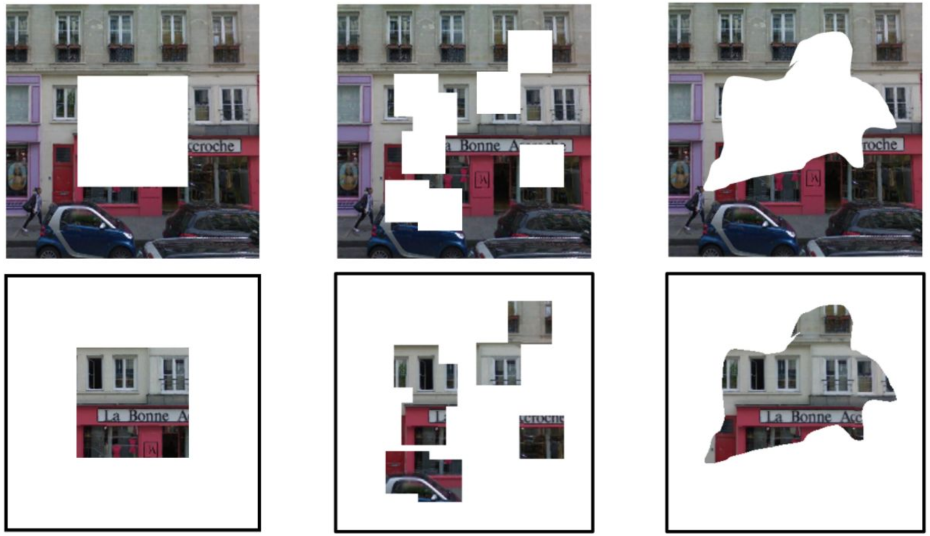
a) Masked Input

b) L2 Loss   c) L2 + adversarial loss

* where x is the ground truth image, F is the context encoder, M is a binary mask corresponding to the dropped image region with a value of 1 wherever a pixel was dropped and 0 for input pixels.

* *Image source:- Deepak Pathak,et al, Context Encoders: Feature Learning by Inpainting. CVPR 2016*

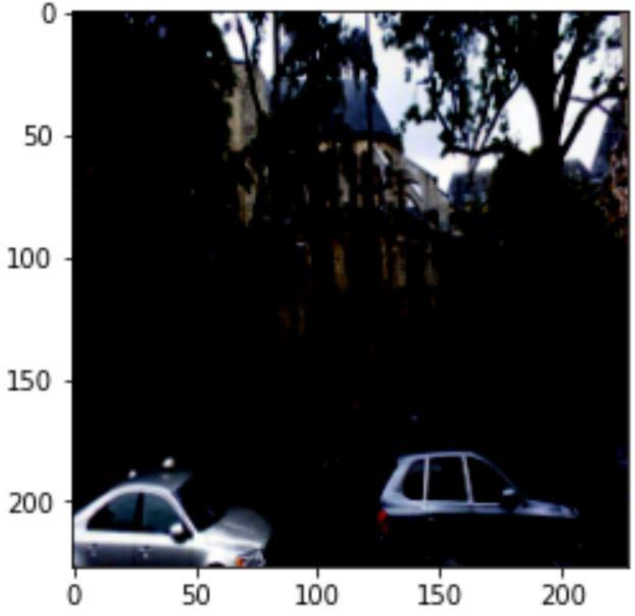
Inpainting - Masks

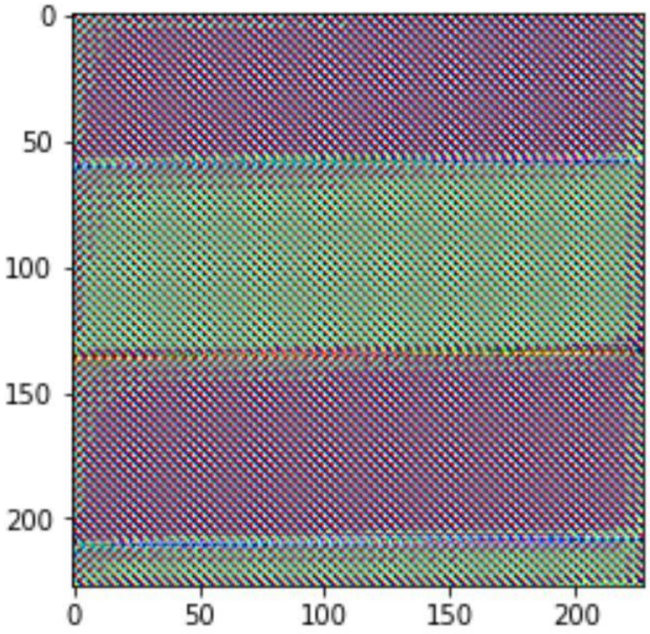


Center region Random blocks Random region

* *Image source:- Deepak Pathak,et al, Context Encoders: Feature Learning by Inpainting. CVPR 2016*

Inpainting - Results





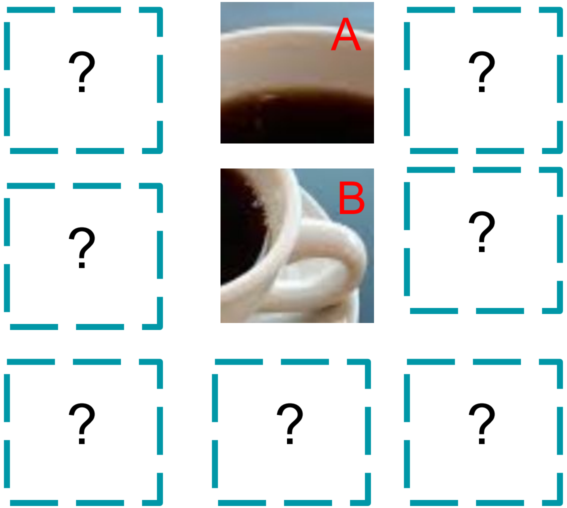
PRETEXT TASK

B. JIGSAW PUZZLES

Inpainting

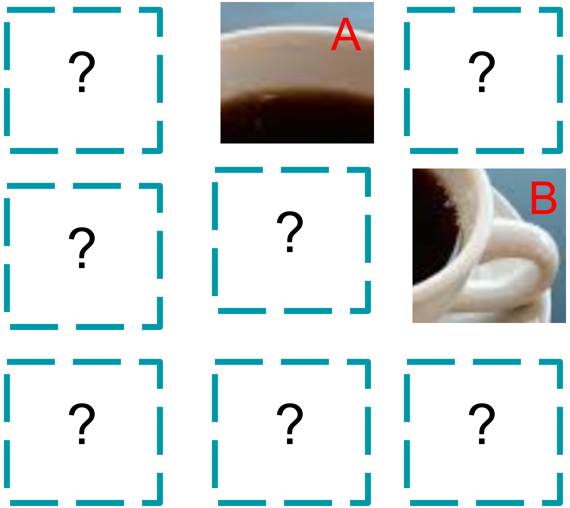
* *Image source:-Google images*

Inpainting



* *Image source:-Google images*

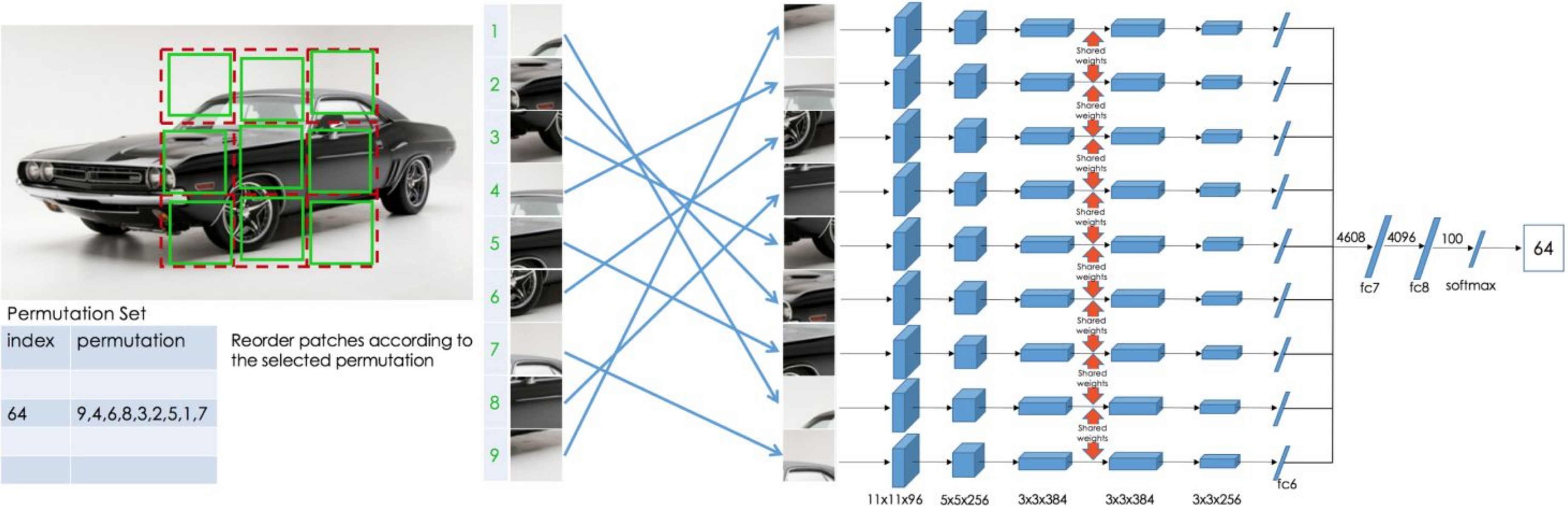
Inpainting



** Image source:-Google images*

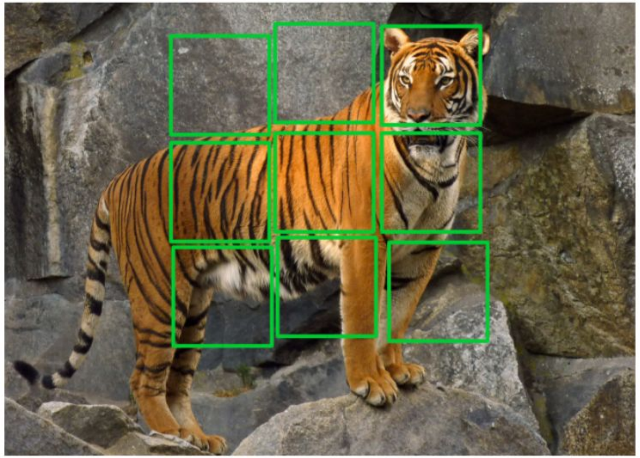
Jigsaw - Implementation

- Siamese network
- Permutations with large Hamming distance



* *Image source:- Noroozi et al - Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles, ECCV 2016*

Jigsaw puzzle



The image from which the tiles (marked with green lines) are extracted

Puzzle obtained by shuffling the tiles

Expected output

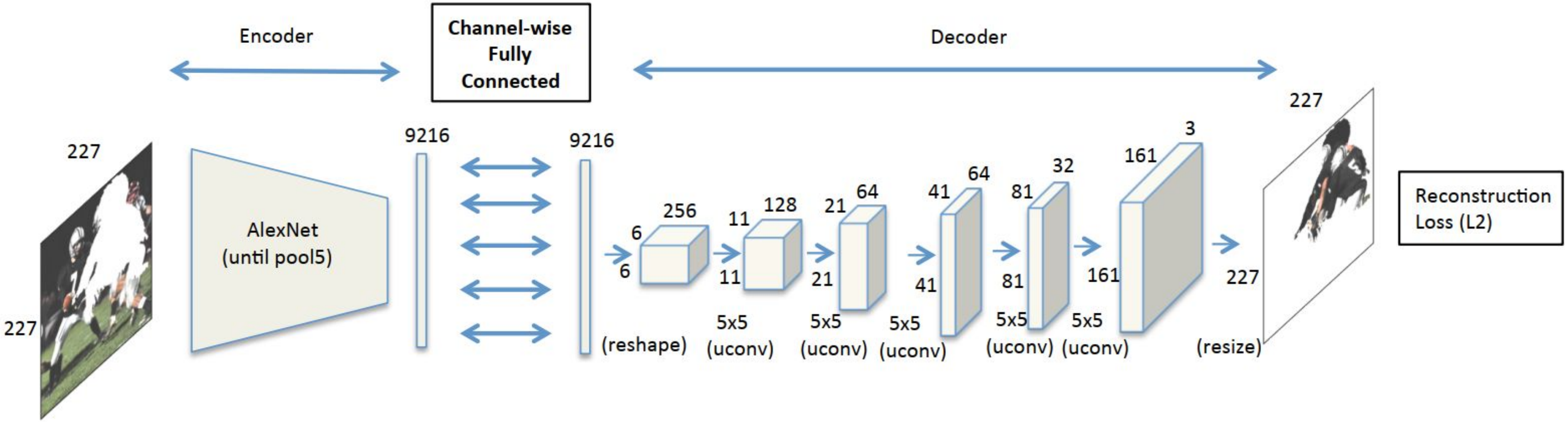
* *Image source:- Noroozi et al - Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles, ECCV 2016*

THANK YOU!

References

1. Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell and Alexei A. Efros, ”Context Encoders: Feature Learning by Inpainting”, CVPR 2016.
2. Noroozi, M. and P. Favaro. “Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles”, ECCV (2016).
3. Carl Doersch, Abhinav Gupta, and Alexei A. Efros. “Unsupervised Visual Representation Learning by Context Prediction”, ICCV 2015
4. Carl Doersch ICCV presentation - http://videolectures.net/iccv2015_doersch_visual_representation/

Inpainting - Architecture



Context encoder trained with reconstruction loss for feature learning by filling in *arbitrary region dropouts* in the input.

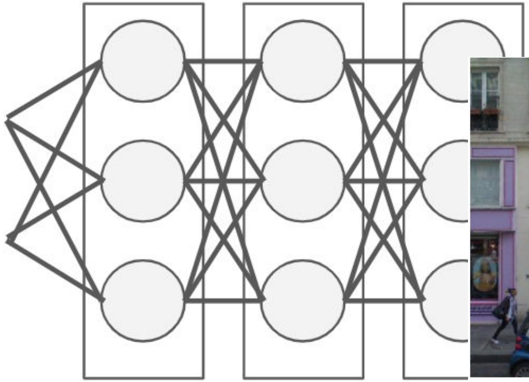
* *Image source:- Deepak Pathak,et al, Context Encoders: Feature Learning by Inpainting. CVPR 2016*

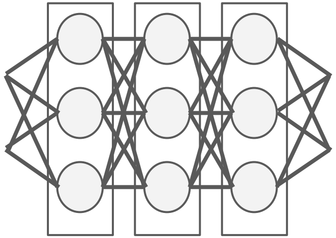
Motivation

1) In motivation add label intensive data is not available
2) Add how context helps in classification(beagle example from carl video)
3) How do these features get used later
4) Inpaintin
    a) Encoder decoder diagram(shud make)
    b)
    c) maybe results(last slide)
5) Jigsaw
    a) Siamese network(why it s needed n why it works)
    b)
6)

1. What is Self-Supervised Learning
2. Inpainting
3. Jigsaw
4.

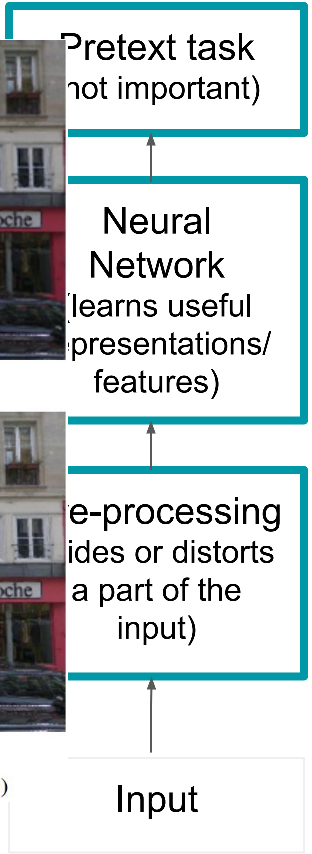
Outli

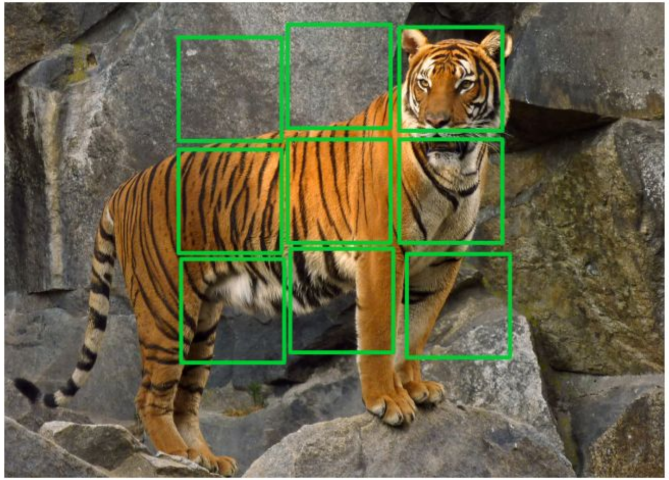




(a) Input context (b) Human artist

(c) Context Encoder ($L2$ loss) (d) Context Encoder ($L2$ + Adversarial loss)





The image from which the tiles (marked with green lines) are extracted

Puzzle obtained by shuffling the tiles

Expected output

Image source:- Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles - Noroozi et al