Improving Landmark Localization with Semi-Supervised Learning

Sina Honari*, Pavlo Molchanov, Stephen Tyree, Pascal Vincent, Christopher Pal, Jan Kautz
University of Montreal, NVIDIA, Ecole Polytechnique of Montreal, CIFAR, Facebook AI Research.

Presented by:
Arif Anjum

All images are taken from the paper Images : https://arxiv.org/abs/1709.01591

Motivation

- *Manual landmark localization is a time consuming and tedious task*
- *To build a database it requires a lot of efforts*
- *Single image labelling requires ~60 sec*
- *But with attributes such as smiling and head orientation(looking straight) image labelling requires ~ 1 sec*

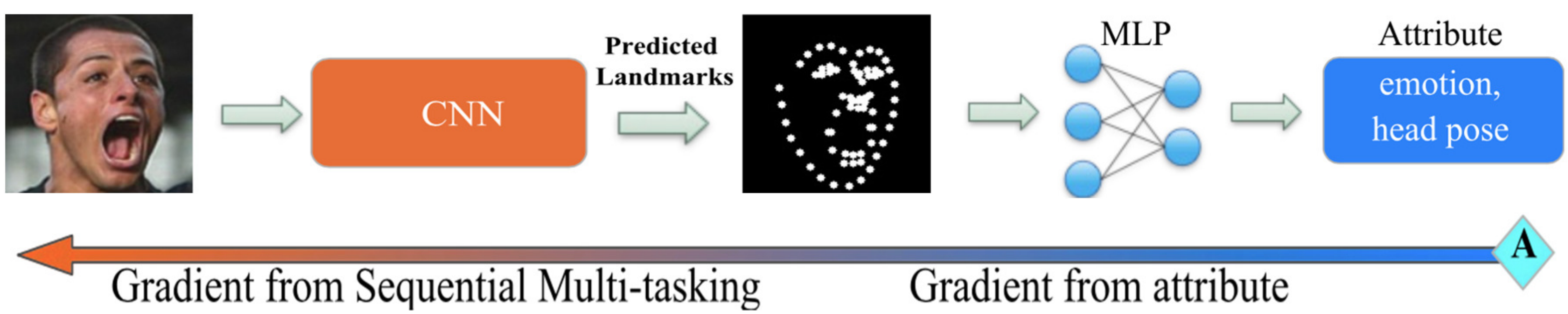
Facial Landmarks Localization

http://mmlab.ie.cuhk.edu.hk/projects/TCDCN.html

Semi-Supervised Learning

Using CNNs with sequential multitasking:

- *Predict Landmarks using Convolutional Neural Network*
- *Use predicted landmarks to predict Attributes*
- *Using backpropagation get the gradient from attributes to the landmark localization network*

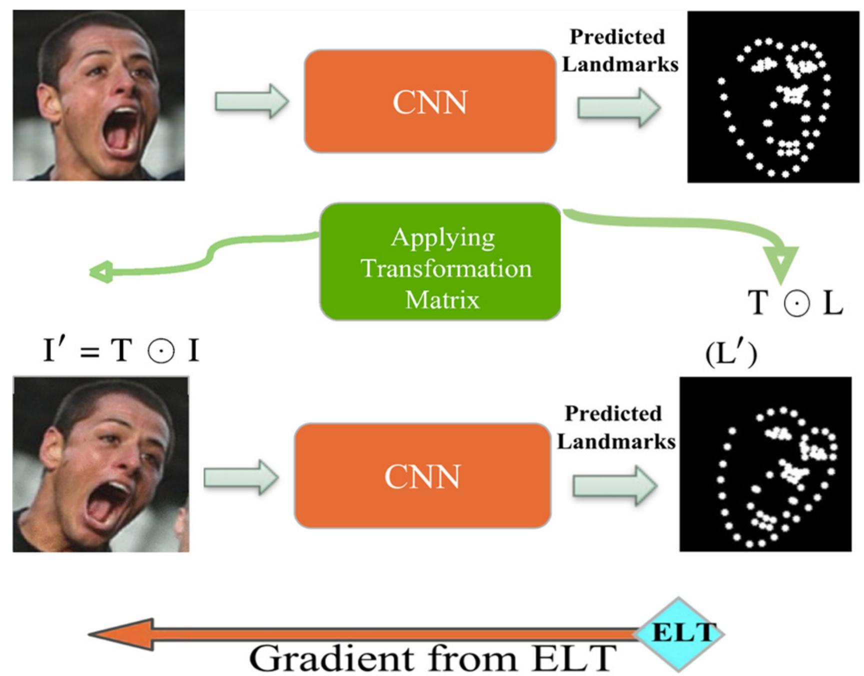


CNN Architecture using Sequential Multi-Tasking

Semi-Supervised Learning

Equivariant Landmark Transformation (ELT):

- *Predict landmarks (L) on an image I*
- *Apply a transformation T to image I*
- *Predict landmark (L') on image I'*
- *Apply transformation T to landmarks L(I)*
- *Compare L' with T ʘ L(I)*
- *T ʘ L(I) ~ L(T ʘ I)*
- *Get gradient from ELT loss:*
  - $ELT = \| T \odot L - L' \|_2^2$

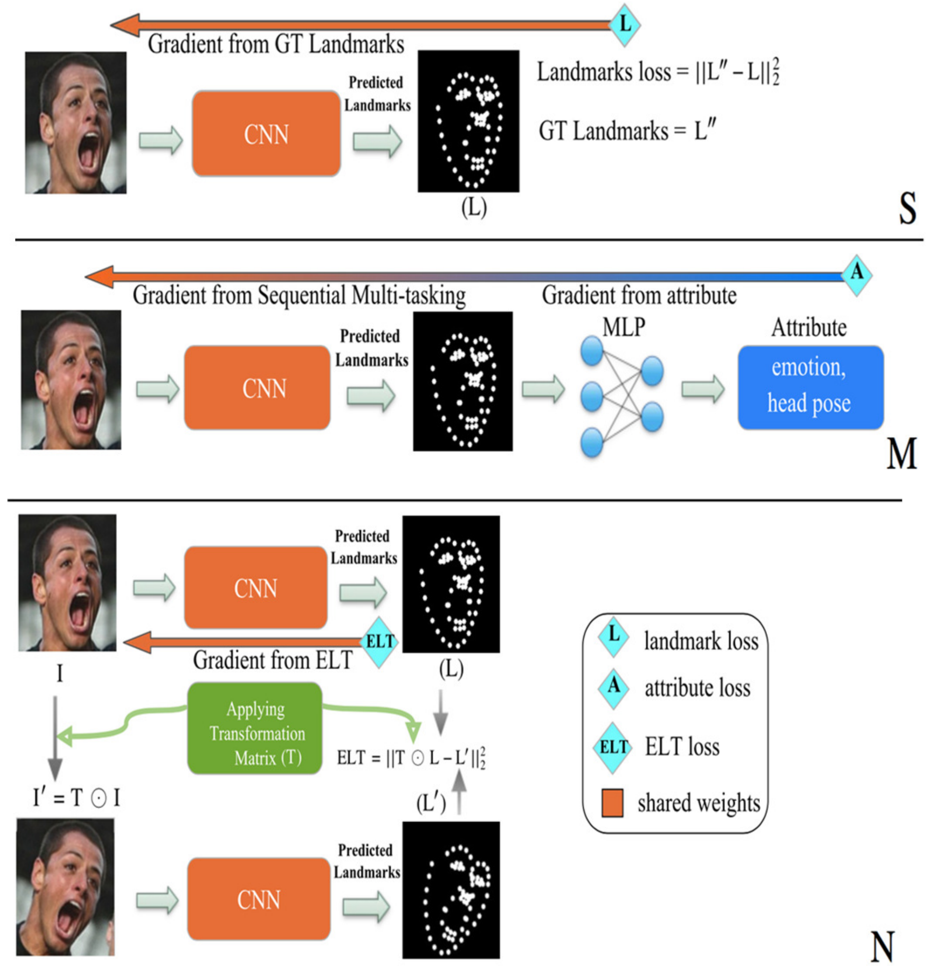


Equivariant Landmark Transformation (ELT):

Learning Landmarks

Making use of all data:

- *Loss from GT Landmarks (L)*
- *Loss from Attributes (A) using Sequential Multi-tasking*
- *Loss from Equivariant Landmark Transformation (ELT)*

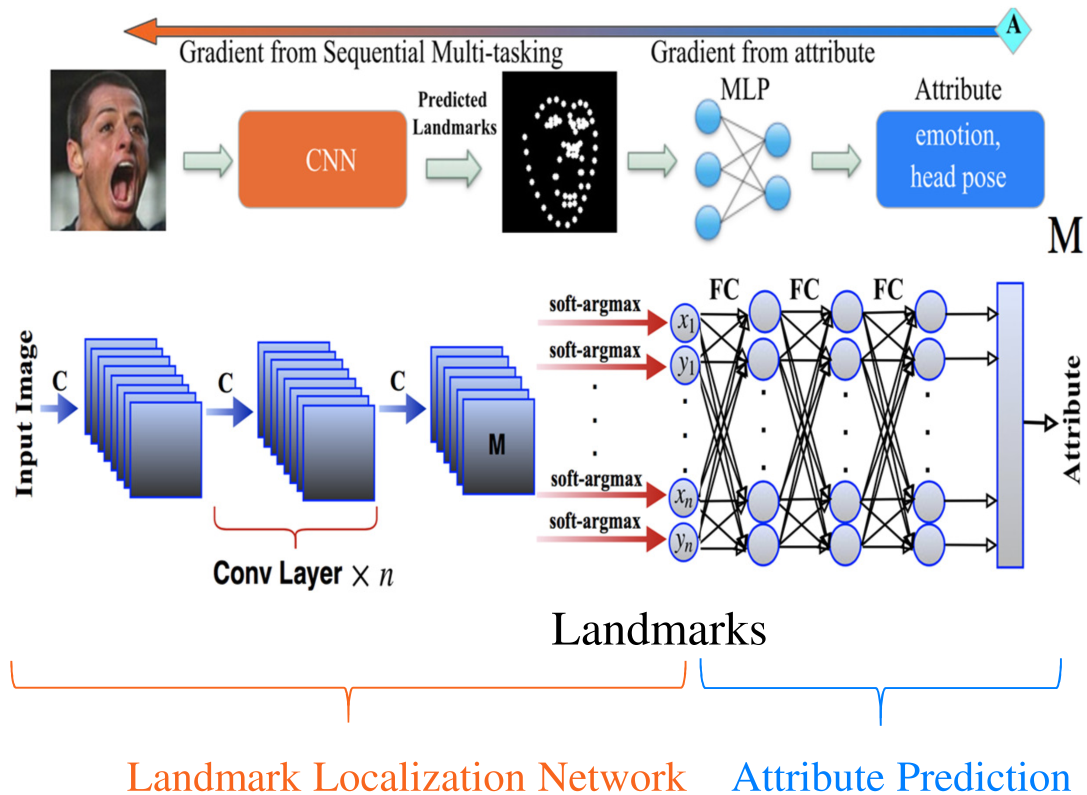


Losses = L + A + ELT

Sequential Multi-Tasking (Seq-MT)

Our basic implementation:

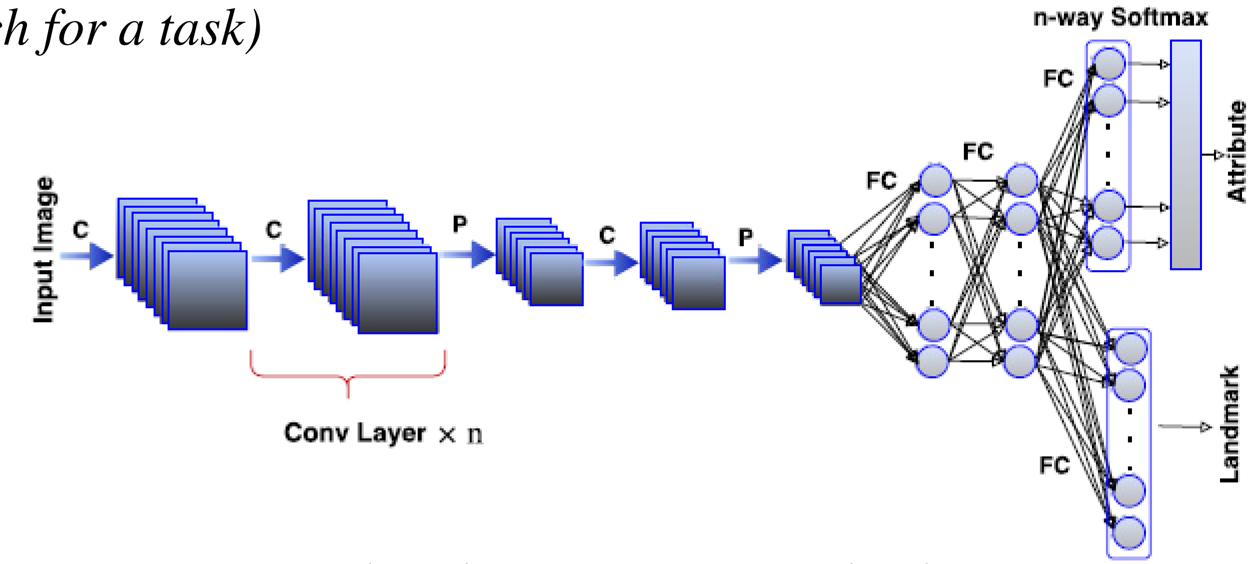
- *Using only conv layers for landmark localization*
- *Predict Attribute only from Landmarks*
- *Using soft-argmax to allow full back-propagation*



CNN Architecture using Sequential Multi-Tasking

Common Multi-tasking (Comm-MT)

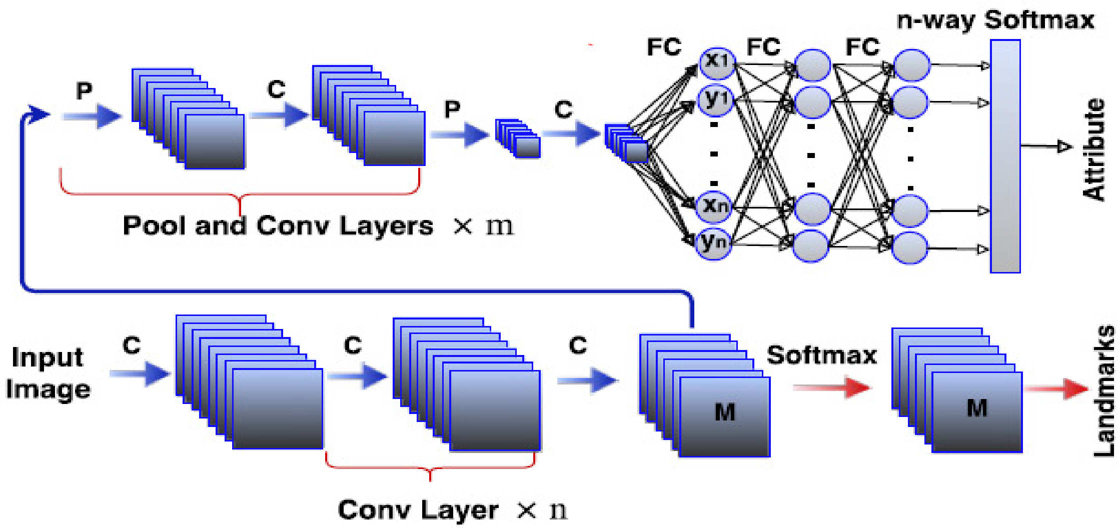
- *The model takes an image and applies a series of conv(C) and pooling (P) layers*
- *Then passed to common (shared) fully connected (FC) layer*
- *The last FC layer connected to two branches (each for a task)*



Common Multi-Tasking (Common -MT) Architechture

Heatmap Multi-Tasking(Heatmap-MT)

- *Landmarks are detected using conv (C) layers without sub-sampling, pooling, or FC layers*
- *A softmax layer is used for landmark prediction in the output layer*
- *Landmark heatmaps right before softmax layer are fed to a series of pool (P) and conv (C) layers which are then passed to FC layers*
- *The last FC layer is fed to softmax for attribute classification*

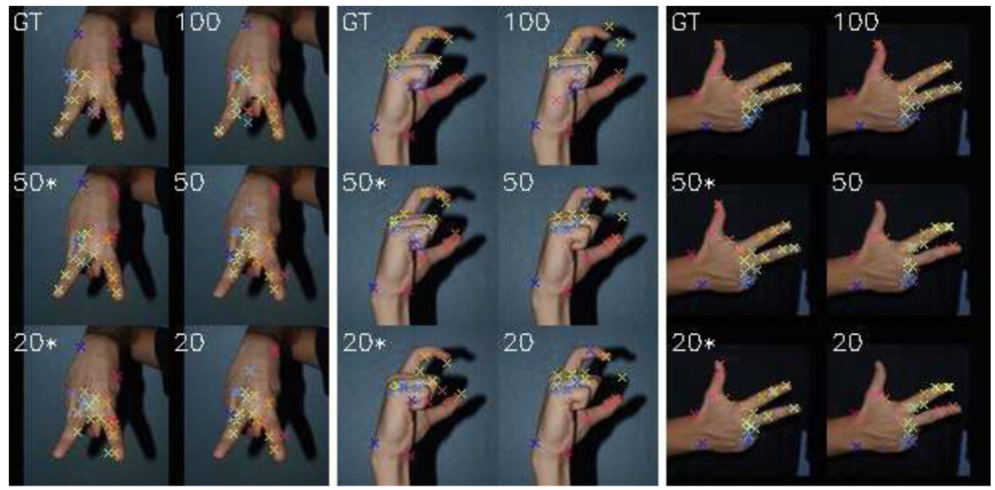


Heatmap Multi Tasking (Heatmap- MT) Architecture

Semi-Supervised Impact on HGR1 hands dataset

Model trained with different percentage of labelled landmarks:

- *Seq-MT loss (L+A) improves results compared to only using landmarks (L)*
- *ELT loss (L+ELT) improves results compared to only using landmarks (L)*
- *Seq-MT (L+ELT+A) compared to Seq-MT (L) gets the same performance with half landmark labels (see 5%, 10%, 20%)*



Model	Percentage of Images with Labeled Landmarks				
	5%	10%	20%	50%	100%
Seq-MT (L)	57.6	41.1	32.0	21.4	**15.8**
Seq-MT (L+A)	50.0	38.1	28.1	19.8	16.9
Seq-MT (L+ELT)	43.7	31.5	25.1	**17.7**	
Seq-MT (L+ELT+A)	**38.5**	**30.3**	**24.0**	19.1	
Comm-MT (L)	77.1	62.8	52.7	41.8	35.7
Comm-MT (L+A)	53.4	39.3	35.5	26.9	24.1
Heatmap-MT (L)	66.5	51.9	42.4	30.9	25.5
Heatmap-MT (L+A)	64.8	54.9	43.2	30.5	26.7

Model predictions on the test set of the HGR1 dataset

Visualizing Results (Multi-PIE)

Training Data	Method	Faces
100 %	*Just CNN*	
5%	*Just CNN*	
5%	*Semi Supervised CNN*	

Model predictions on Multi-PIE Dataset