A Unified Neural Based Model For Structured Output Problems

Soufiane Belharbi, Clément Chatelain, Romain Hérault, Sébastien Adam

1 Context and Motivation: Structured Output Problems (SOP)

Standard Machine Learning

$\mathcal{M}_\theta : \mathcal{X} \to y$

- Inputs $\mathcal{X} \in \mathbb{R}^d$
- Output $y \in \mathbb{R}$: classification, regression, ...

Machine Learning for Structured Output Problems

$\mathcal{M}_\theta : \mathcal{X} \to \mathcal{Y}$

- Inputs $\mathcal{X} \in \mathbb{R}^d$
- Outputs $\mathcal{Y} \in \mathbb{R}^{d'}, d' > 1$ with *structured dependencies*

Motivation

- A priori knowledge about the structure of the outputs helps the prediction
- How to *learn* the structure of the outputs during the training of $\mathcal{M}_\theta$?

2 Our Approach: Deep Neural Networks for SOP (New Formulation)

New Formulation for SOP

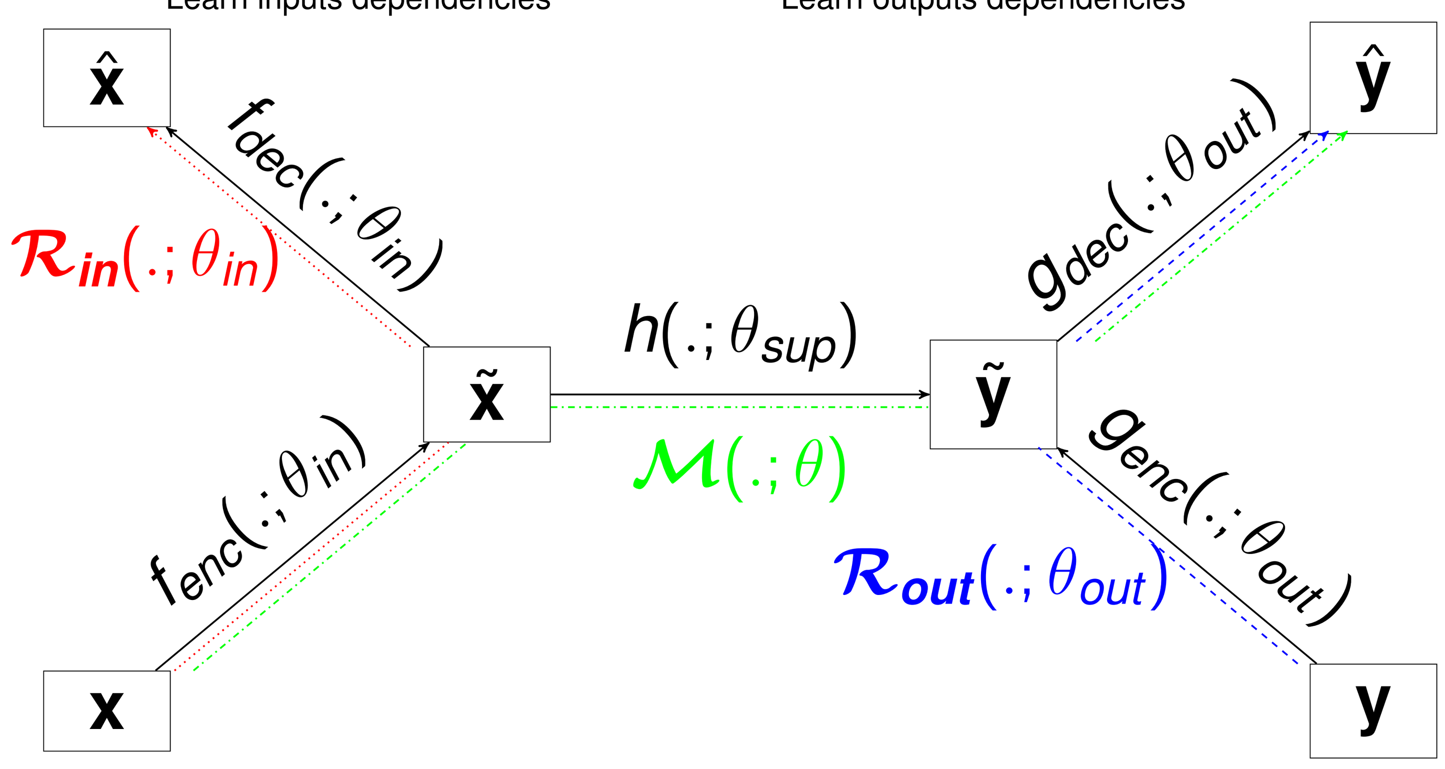
$$\mathcal{L}(\theta, \mathcal{D}(\mathbf{x},\mathbf{y})) = \frac{1}{n}\sum_{i=1}^{n}\Big[\underbrace{\mathcal{C}(\mathcal{M}(x_i;\theta_{sup},\theta_{in},\theta_{out}), y_i)}_{\text{Learn (inputs} \to \text{outputs) dependencies}} + \underbrace{\ell_{in}(\mathcal{R}_{in}(x_i;\theta_{in}), x_i)}_{\text{Learn inputs dependencies}} + \underbrace{\ell_{out}(\mathcal{R}_{out}(y_i;\theta_{out}), y_i)}_{\text{Learn outputs dependencies}}\Big] \quad (1)$$

Our Approach

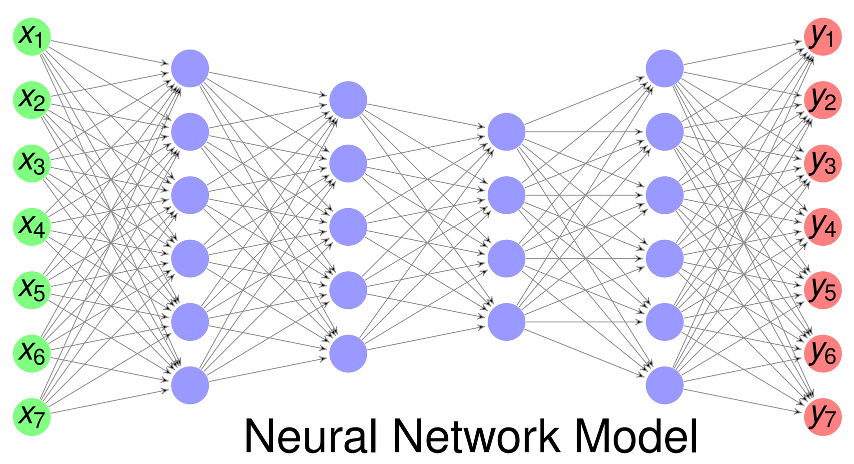
- Based on Deep Neural Network (IODA)[1]
- Use the layer-wise pre-training trick to:
  - Learn inputs dependencies
  - *Learn outputs dependencies*
- Pre-trainer: Auto-encoder

$\mathcal{C}(.), \ell_{in}(.), \ell_{out}(.)$: defined costs.

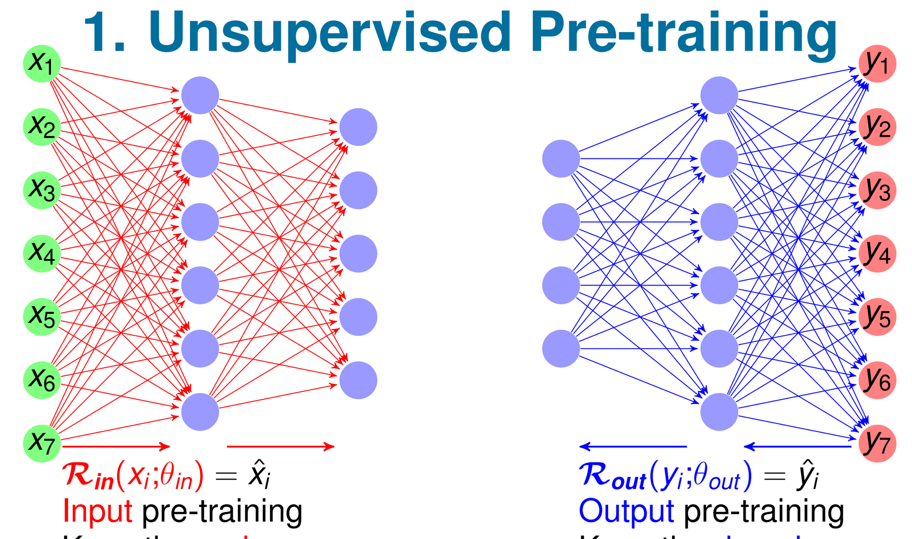
$\mathcal{R}_{in}(.)/\mathcal{R}_{out}(.)$: Input/Output reconstruction. $\mathcal{M}(.)$: supervised task.



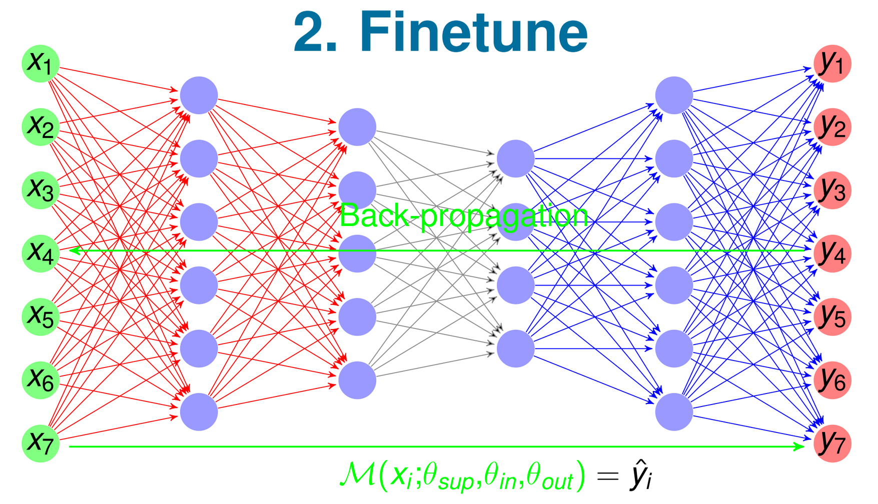
3 Training



1. Unsupervised Pre-training



2. Finetune



4 Application: Facial Landmark Detection (Regression Approach)

Facial Landmark Detection Problem

NRMSE:

Normalized **R**oot **M**ean **S**quare **E**rror between the predicted shape and ground truth shape.

CDF$_{0.1}$:

Cumulative **D**istribution **F**unction (percentage of test images that have an NRMSE less or equal than 0.1).

AUC:

Area **U**nder the CDF_x **C**urve.

	LFPW		HELEN	
	AUC	$CDF_{0.1}$	AUC	$CDF_{0.1}$
Mean shape	66.15%	18.30%	63.30%	16.97%
NDA 0-0-0	77.60%	50.89%	80.91%	69.69%
IDA 1-0-0	79.25%	62.94%	82.13%	76.36%
2-0-0	79.10%	58.48%	82.39%	75.75%
3-0-0	79.51%	65.62%	82.25%	77.27%
IODA 1-0-1	80.66%	68.30%	83.95%	83.03%
1-1-1	81.50%	72.32%	83.51%	80.90%
1-0-2	81.00%	71.42%	83.91%	82.42%
1-1-2	81.06%	70.98%	83.81%	83.03%
1-0-3	**81.91%**	**74.55%**	83.72%	80.30%
2-0-1	81.32%	72.76%	83.61%	80.00%
2-1-1	81.47%	70.08%	**84.11%**	**83.33%**
2-0-2	81.35%	71.87%	83.88%	82.12%
3-0-1	81.62%	72.76%	83.38%	78.48%

Table : Performance of: mean shape, **NDA**, **IDA** and **IODA** on LFPW and HELEN datasets.

NDA: **N**on pre-trained **D**eep **A**rchitecture
IDA: **I**nput pre-trained **D**eep **A**rchitecture
IODA: **I**nput/**O**utput pre-trained **D**eep **A**rchitecture

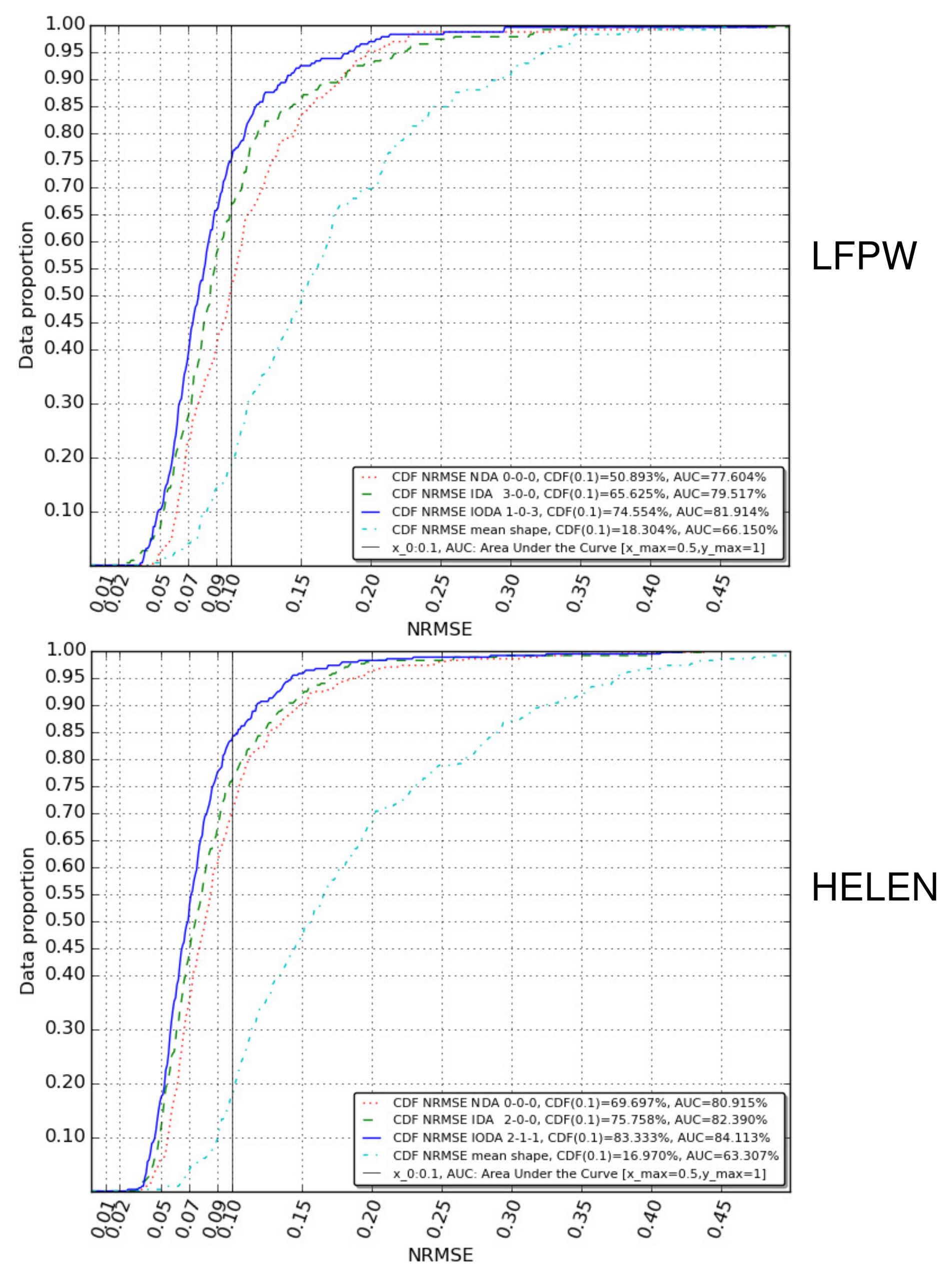


Figure : CDF_x curves of best configurations

5 Perspectives

- Minimize Eq.1 at the same time using weighted sub-losses

6 References

[1]:J. Lerouge, R. Herault, C. Chatelain, F. Jardin, and R. Modzelewski. IODA: An Input Output Deep Architecture For Image Labeling. Pattern Recognition, 48(9):2847-2858, 2015.

Conférence d'APprentissage automatique (CAP), 2015, Lille soufiane.belharbi@insa-rouen.fr, projet ANR JCJC Lemon, www.litislab.eu