Tencent AI Lab

City U 香港城市大學 City University of Hong Kong

NDDR-CNN: Layerwise Feature Fusing in Multi-Task CNNs by Neural Discriminative Dimensionality Reduction

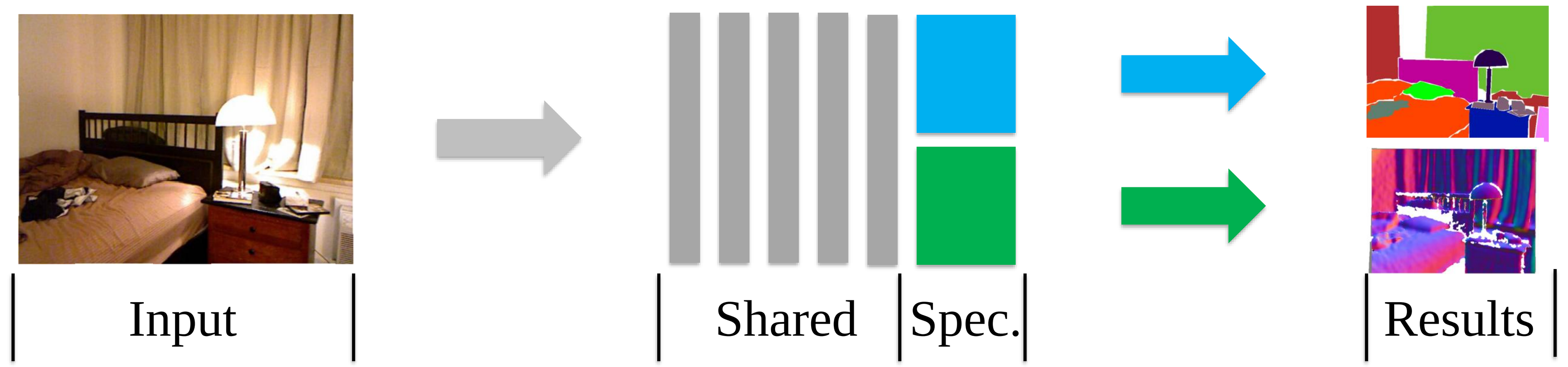
Yuan Gao, Jiayi Ma, Mingbo Zhao, Wei Liu, Alan Yuille

JOHNS HOPKINS UNIVERSITY

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

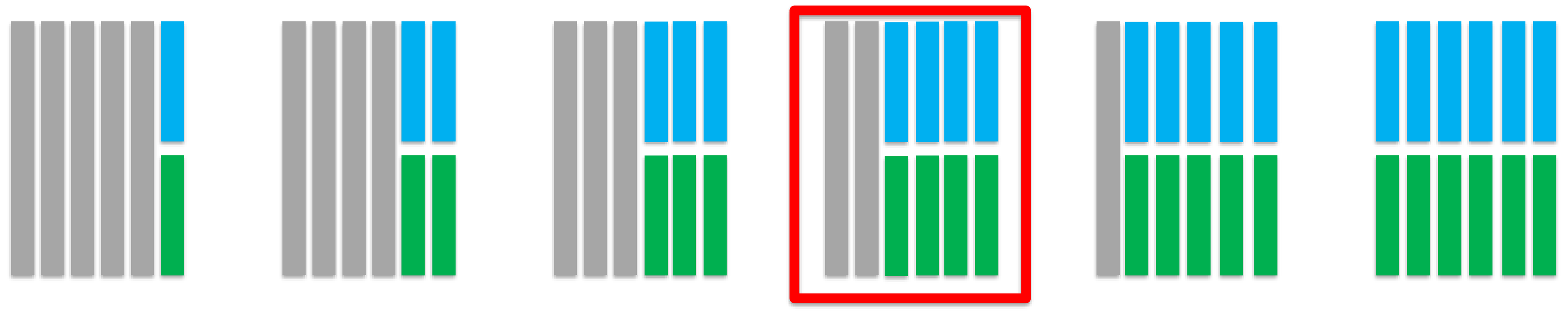
Motivation

Standard Multi-Task CNNs:



Q: Is it appropriate to assume that the low- and mid-level features for different tasks in MTL should be identical (i.e., shared), and only split at the last layer?

A: Empirical splitting/sharing leads to different results [1]!



Best arch can be ***Task/Dataset-Related*** and achieves in ***Middle***!

Infeasible Exhaust Search

Generalization Problem

Improper arch harms to one or more tasks

→ *General Purpose*, *Efficient*, *Plug and Play*, Multi-task Learning CNN architecture

Key Ideas

1. Start with ***Single Task Networks***.
2. Explore ***feature embedding*** from ***different tasks*** at ***every CNN level***.

This can be achieved by a novel ***Neural Discriminative (supervised) Dimensionality Reduction (NDDR)*** operation formulated by ***1x1 Conv***, ***BatchNorm***, and ***Weight Decay***.

Methodology

Equivalence of NDDR and Concat + 1x1 Conv + BatchNorm + WeightDecay:

1. $\forall$ task i, concat features $F_l^i \in \mathbb{R}^{N\times H\times W\times C}$ from layer l:

$$\mathbf{F}_l = \textsc{Concat}[\mathbf{F}_l^1, ..., \mathbf{F}_l^K] \in \mathbb{R}^{N\times H\times W\times KC}$$

2. Perform feature projection on the concated features:

$$\mathbf{F}_l^{i*} = \mathbf{F}_l\mathbf{W} \in \mathbb{R}^{N\times H\times W\times C}$$

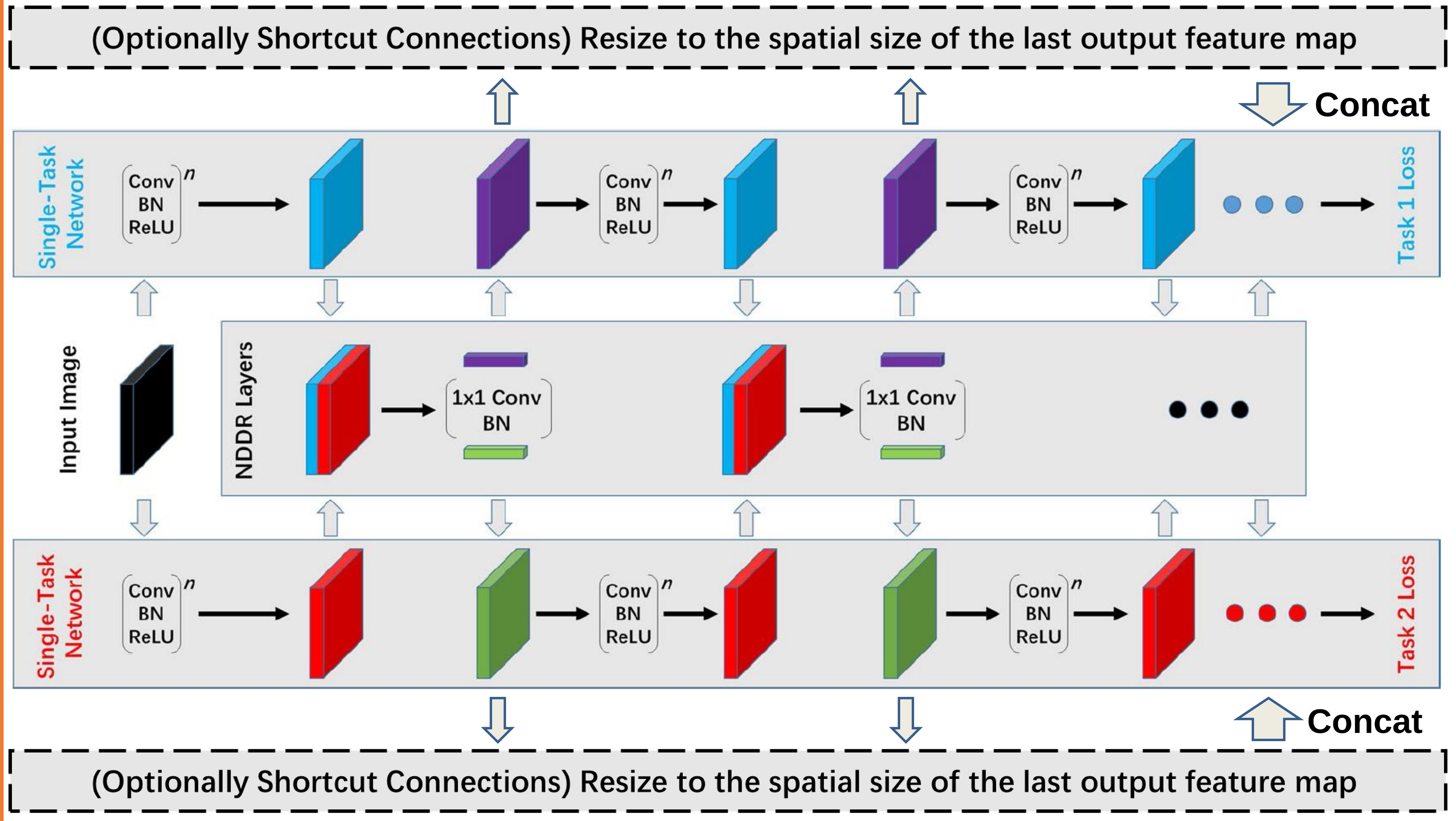
1x1 Conv with fea_in KC, fea_out C!

$$\mathbf{W} \in \mathbb{R}^{KC\times C} \xrightarrow{\text{reshape}} \mathbf{W} \in \mathbb{R}^{KC\times 1\times 1\times C}$$

3. Constraints on Transformation $\mathbf{W}$ ⟶ **Weight Decay**

Normalization on Input $\mathbf{F}_l^{i*}$ ⟶ **BatchNorm**

Overview of our networks:



Features

1. ***General Purpose:*** w/o tasks or dataset assumption;
2. ***Efficient:*** very few additional NDDR layers enables significant boost (e.g., 5 vs. 101 in ResNet-101);
3. ***Plug and Play:*** support most (if not all) existing CNN arches in an end-to-end trainable manner.
4. ***Robust to Hyperparas:*** see the ablations on weight init., learning rate init., and pretrains in our paper!

Experiments

Pixel Labeling Tasks: Seg. + Surface Normal

	Surface Normal Prediction					Semantic Seg.	
	Errors (Lower Better)		Within t° (%) (Higher Better)			(%) (Higher Better)	
	Mean	Med.	11.25	22.5	30	mIoU	PAcc
Sing.	15.6	12.7	44.3	74.8	87.2	39.5	69.2
Mult.	16.3	13.8	41.1	73.9	86.5	39.1	68.7
C.-S.	15.9	13.2	42.9	75.1	86.8	40.5	69.5
Sluice	15.3	12.8	44.1	76.9	88.2	40.8	70.1
Ours	**14.4**	**11.6**	**48.5**	**79.1**	**89.5**	**43.3**	**71.5**

Results w/ ResNet-101-Deeplab

	Surface Normal Prediction					Semantic Seg.	
	Errors (Lower Better)		Within t° (%) (Higher Better)			(%) (Higher Better)	
	Mean	Med.	11.25	22.5	30	mIoU	PAcc
Sing.	15.4	12.1	46.9	76.1	86.9	33.5	64.4
Mult.	15.2	11.8	48.0	76.4	87.0	33.1	64.0
C.-S.	14.8	11.1	50.3	76.9	87.0	35.0	65.1
Sluice	14.2	10.6	51.7	78.2	88.2	35.3	65.3
Ours	**13.4**	**9.8**	**55.1**	**80.5**	**89.4**	**36.7**	**66.9**

Results w/ VGG16-Deeplab-Shortcut

Image Level Tasks: Age + Gender Est.

	Age (Lower Better)		Gender (Higher Better)
	Mean AE	Median AE	Acc. (%)
Single-Task	9.1	7.4	83.5
Multi-Task	9.0	7.4	82.3
Cross-Stitch	8.6	7.0	**84.0**
Sluice	8.5	7.0	84.1
Ours	**8.0**	**6.2**	**84.0**

Results w/ VGG16

Our Code is Released!

https://github.com/ethanygao/NDDR-CNN

Reference

1. Misra, Shrivastava, Gupta, Hebert. CVPR 2016.
2. Ruder, Bingel, Augenstein, Søgaard, AAAI 2019.
3. Chen, Papandreou, Kokkinos, Murphy, Yuille. TPAMI 2018.
4. He, Zhang, Ren, Sun. CVPR2016. 5.Simonyan, Zisserman. ICLR2015.