Learning Unified Representations for Multi-Resolution Face Recognition

Hulingxiao He[1], Wu Yuan[1], Yidian Huang[1], Shilong Zhao[1], Wen Yuan*[2], Hanqing Li[2]

[1] Beijing Institute of Technology [2] Institute of Geographic Sciences and Natural Resources Research, CAS

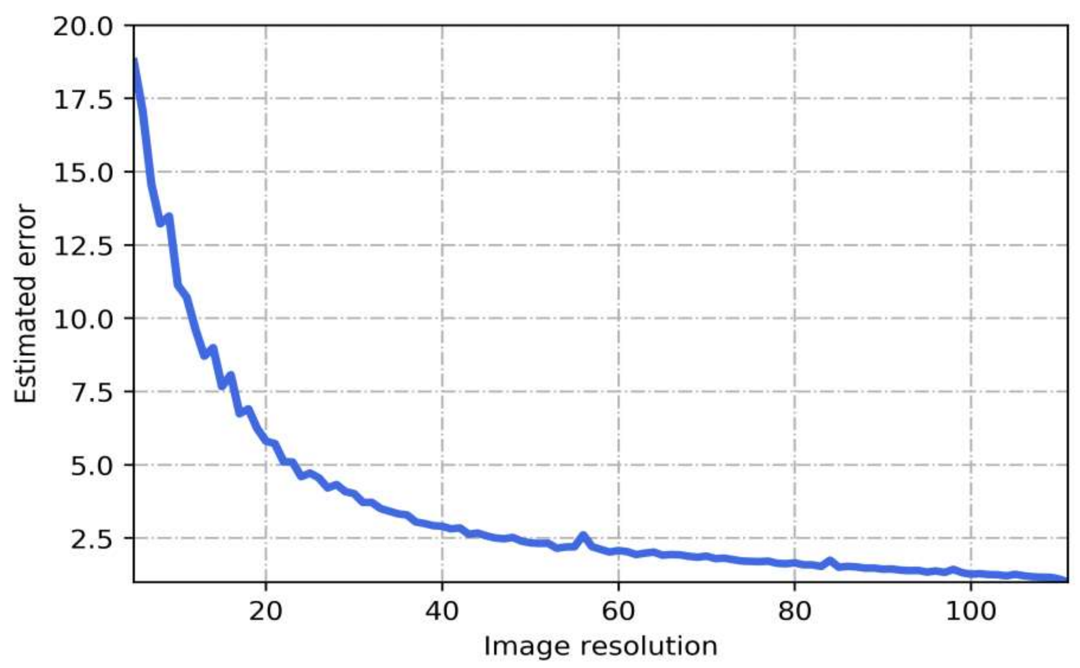
Paper: https://arxiv.org/pdf/2310.09563.pdf **Code:** https://github.com/StevenSmith2000/BTNet

BMVC 2023

Motivation

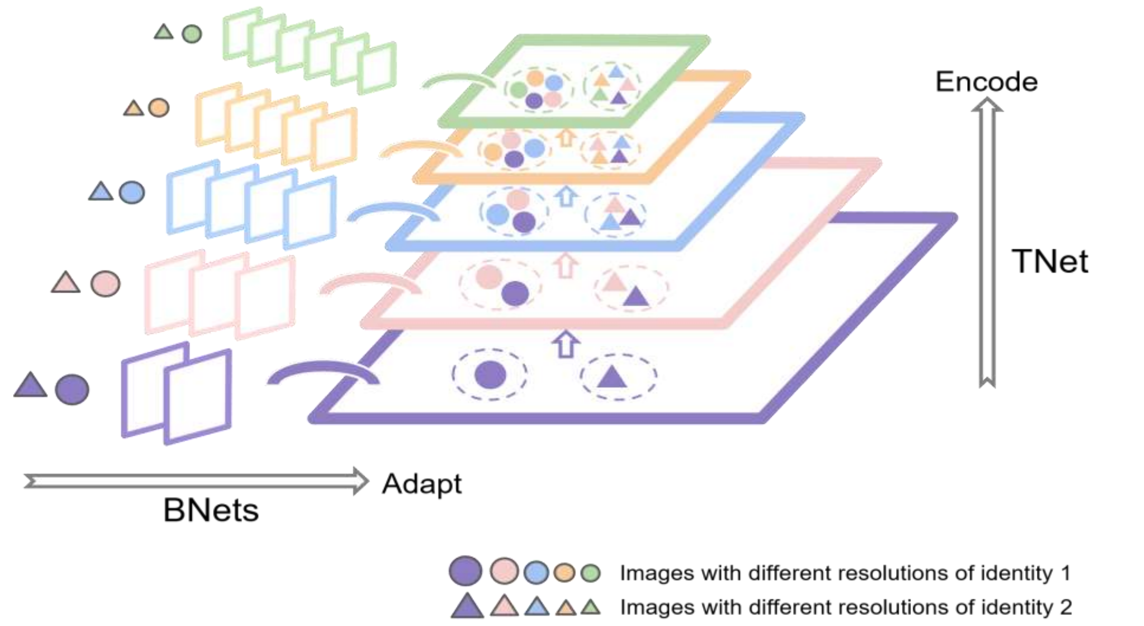
The resolutions of face images in reality may be far beyond the scope covered by the model. As the small feature maps with a fixed spatial extent (e.g., 7×7) are mapped to an embedding with a predefined dimension (e.g., 128d) by a fully connected layer, input images need to be rescaled to a canonical spatial size (e.g., 112 × 112) before fed into the network.

The figure illustrates the experimental estimation of **interpolation error**, whose upper bound increases with the decline of the image resolution.



Overview

We propose a novel representation learning method called **Branch-to-Trunk Network** for multi-resolution face recognition. The method uses a trunk network (TNet) and multiple branch networks (BNets) to improve the discriminability of tiny faces by mitigating the interpolation error introduced by rescaling. Our method achieves strong performance in face recognition tasks while reducing computation and parameter storage. Related work involves compatible representation learning, knowledge distillation and transfer, and low-resolution face recognition.

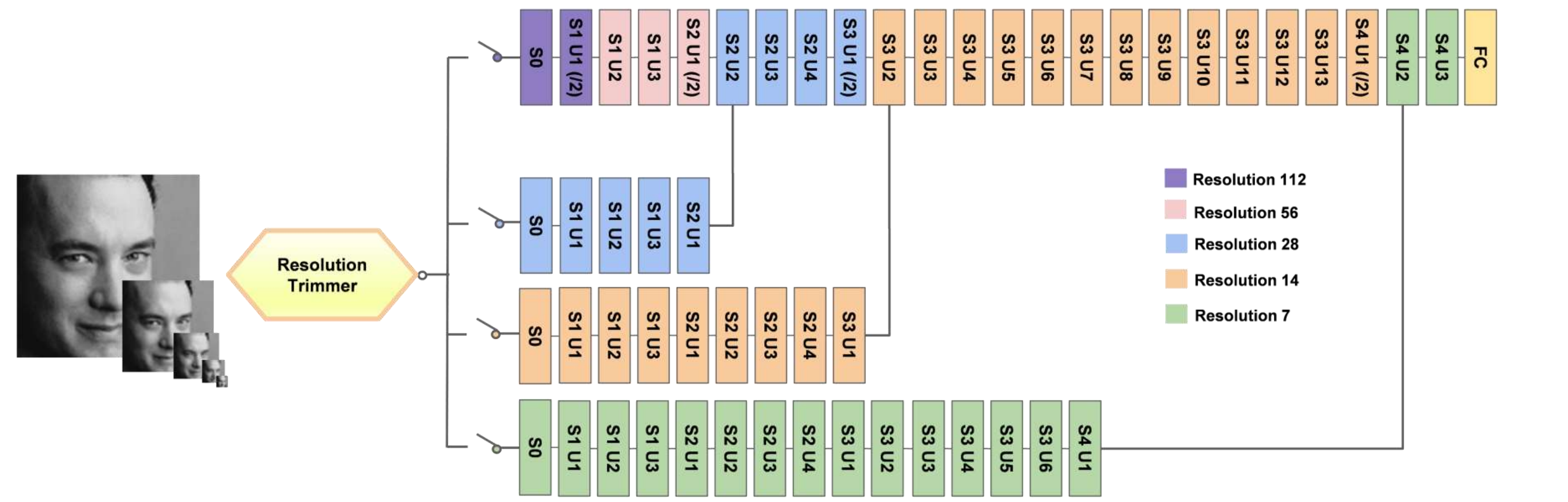


Branch-to-Trunk Network

In BTNet, each input image is assigned a resolution-specific branch (BNets) through a branch selection process. The output of the branch networks is implanted as feature maps in the feature pyramid of the trunk network (TNet). This allows for the transfer of high-resolution information to multiple branches while maintaining representation compatibility.

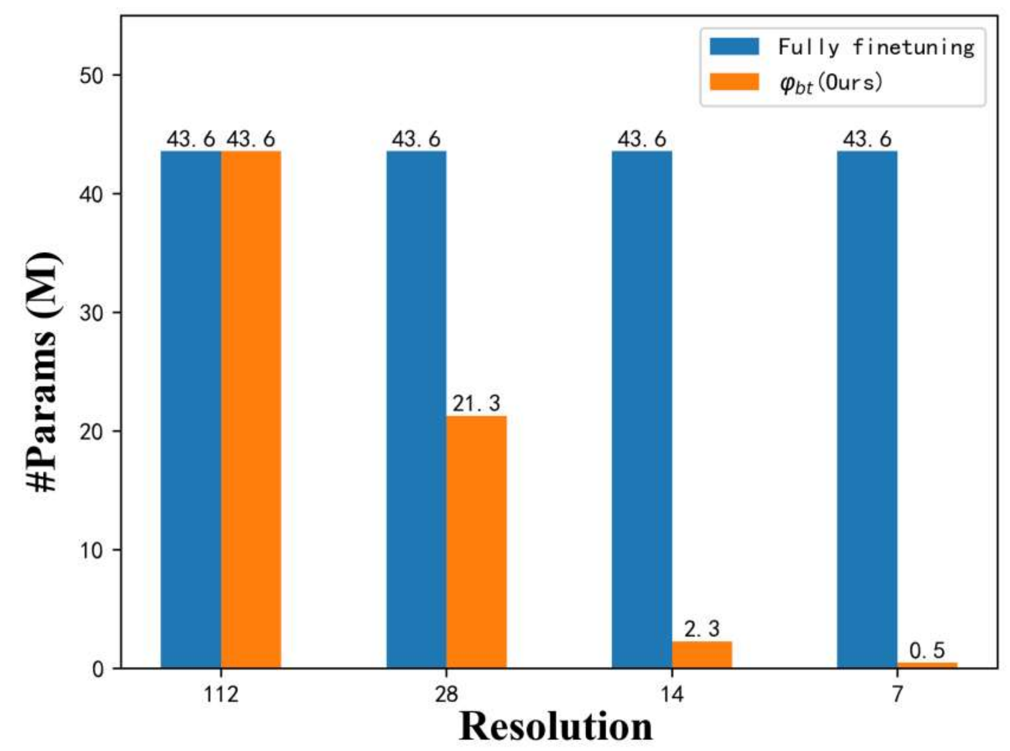
The BTNet architecture is based on a unified encoder (TNet) and resolution adapters (BNets). The BNets focus on resolution-specific feature transfer, while the TNet extracts discriminative information from different resolutions. The overall workflow can be summarized into four steps: **branch selection, resolution adaptation, feature embedding, and classification.**

Detailed architecture of BTNet-res50 (ϕ_{bt}) is below.



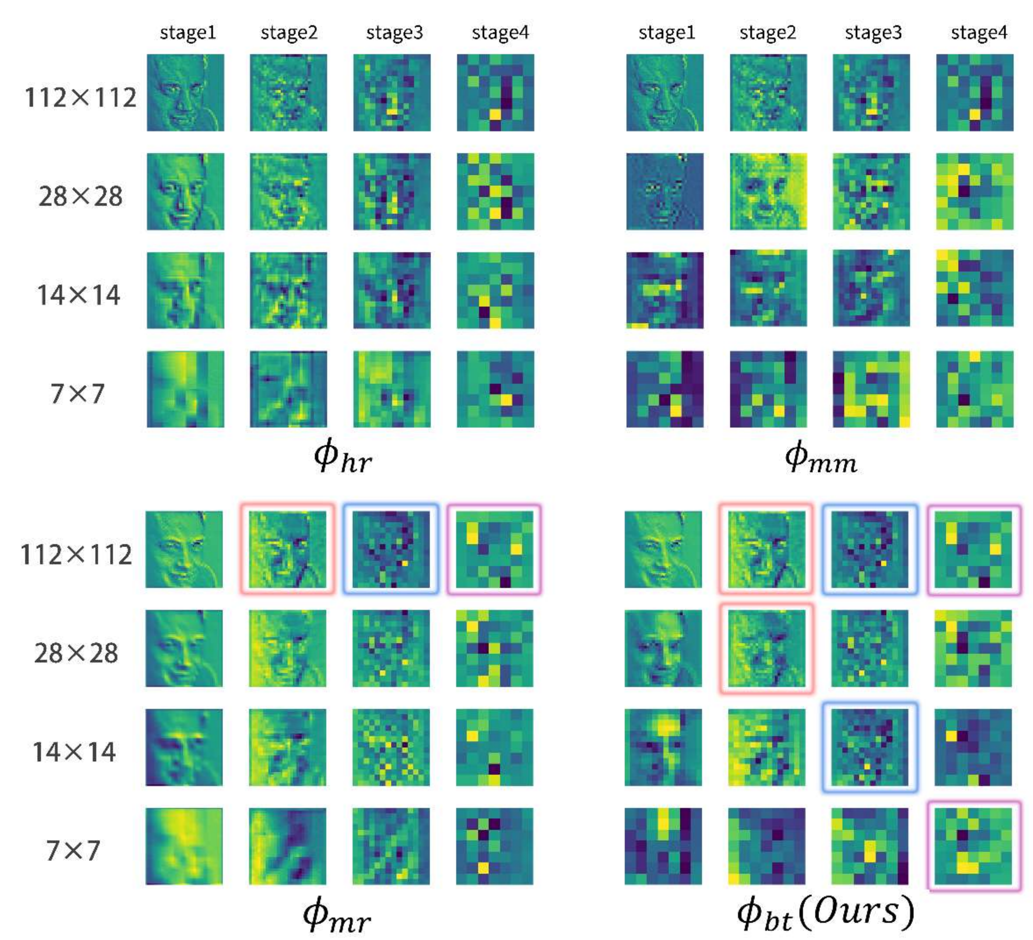
Reducing Parameter Storage

We only need to store the learned branches and re-use the original copy of the pretrained trunk model, significantly reducing the storage cost. BTNet requires only **1.1% ~ 48.9%** of all the parameters compared to fully updating all the parameters of TNet.



Visualization

Through the resolution-specific feature transfer of multiple branches, BTNet can encourage the transferred features to be aligned before fed into the trunk network in corresponding layers.



Experiments & Results

1. Multi-Resolution Face Identification: The results show that BTNet achieves **SOTA performance on QMUL-SurvFace 1:N face identification task**, while also being more computationally efficient.
2. Multi-Resolution Face Verification: Two different settings are considered for multi-resolution identity matching - same-resolution matching and cross-resolution matching. The average performance of BTNet is compared with other models on popular benchmarks, and BTNet consistently outperforms other models.
3. Ablation Study: We analyze the training method alternatives, influence loss implementation alternatives and specific-shared layer allocation alternatives. Results show BTNet have effective training strategies, flexible implementation of loss and better parameter/accuracy tradeoffs.

	TPIR20(%)@FPIR			
	AUC	**0.3**	**0.2**	**0.1**
VGG-Face	14.0	5.1	2.6	0.8
DeepID2	20.8	12.8	8.1	3.4
FaceNet	19.8	12.7	8.1	4.3
SphereFace	28.1	21.3	15.7	8.3
SRCNN	27.0	20.0	14.9	6.2
FSRCNN	27.3	20.0	14.4	6.1
VDSR	27.3	20.1	14.5	6.1
DRRN	27.5	20.3	14.9	6.3
LapSRN	27.4	20.2	14.7	6.3
ArcFace	25.3	18.7	15.1	10.1
RAN	32.3	26.5	21.6	14.9
SST	-	12.4	-	9.7
MASST	-	12.2	-	9.2
MIND-Net	31.9	25.5	-	20.4
AdaFace	32.6	28.3	23.6	16.5
BTNet (avg.+floor)	32.6	27.9	23.4	16.5
BTNet (avg.+near)	34.6	30.3	25.7	18.9
BTNet (avg.+ceil)	**35.4**	31.1	26.8	20.3
BTNet (min+floor)	32.3	27.6	23.2	16.1
BTNet (min+near)	34.0	29.6	25.0	18.0
BTNet (min+ceil)	35.3	31.0	26.6	19.9
BTNet (max+floor)	33.6	29.1	24.5	17.6
BTNet (max+near)	35.2	31.0	26.4	19.6
BTNet (max+ceil)	**35.4**	31.2	26.9	20.6