MULTIGRID NEURAL ARCHITECTURES

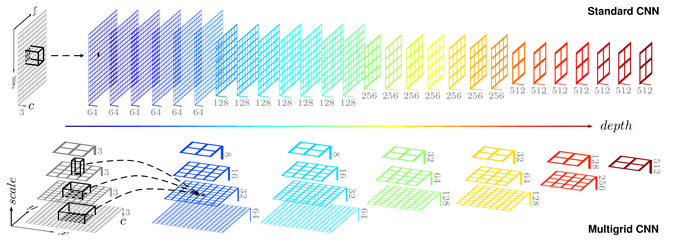
Tsung-Wei Ke — *UC Berkeley / ICSI*

Michael Maire — *TTI-Chicago*

Stella X. Yu — *UC Berkeley / ICSI*

BAIR — BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

TOYOTA TECHNOLOGICAL INSTITUTE AT CHICAGO



Key Ideas:

- Endow CNNs with internal cross-scale communication pathways (*multigrid architecture*)
- Representations live on a pyramid of grids; convolution acts across scale-space

Standard CNNs (since [1])	conflate abstraction & scale	⟶	**disentangle abstraction & scale**
	fine-to-coarse	⟶	**coarse & fine** (optionally progressive)
	slow receptive field growth	⟶	**fast receptive field growth**

Multigrid CNNs

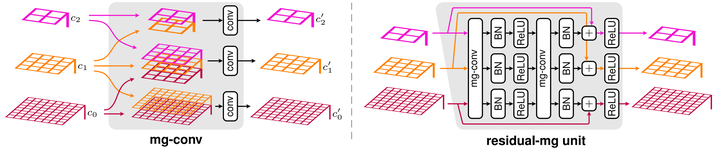
Consequences:

- Receptive field grows exponentially with depth, facilitating rapid integration of context
  → destroys traditional notion of receptive field; short connection paths allow data-dependent routing
- Can learn internal **attention** mechanisms, and accomplish tasks that standard CNNs fail to learn

Experimental Results:

- Classification & segmentation: **bests standard ResNet**; better performance per layer, parameter, & FLOP
- Spatial transformers: can learn task end-to-end; standard CNNs cannot (multigrid handles translation)

MULTIGRID LAYER IMPLEMENTATION



- Multigrid convolution (mg-conv) assembled from standard components:
  upsampling: nearest neighbor, **downsampling:** max-pooling, **concatenation**, & (standard) **convolution**
- Rethink pooling: communication, not summarization → can remove explicit spatial resolution reduction stages
- Residual multigrid networks: apply multigrid convolution, apply batch norm (BN) per grid; see (residual-mg unit)

[1] K. Fukushima. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 1980.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. *CVPR*, 2016.

[3] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial Transformer Networks. *NIPS*, 2015.

[4] M. Maire, and S. Yu. Progressive Multigrid Eigensolvers for Multiscale Spectral Segmentation. *ICCV*, 2013.

[5] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, 2015.

[6] Z. Sergey, and K. Nikos. Wide Residual Networks. *BMVC*, 2016.

[7] K. Simonyan, and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR*, 2015.

NETWORK DIAGRAMS

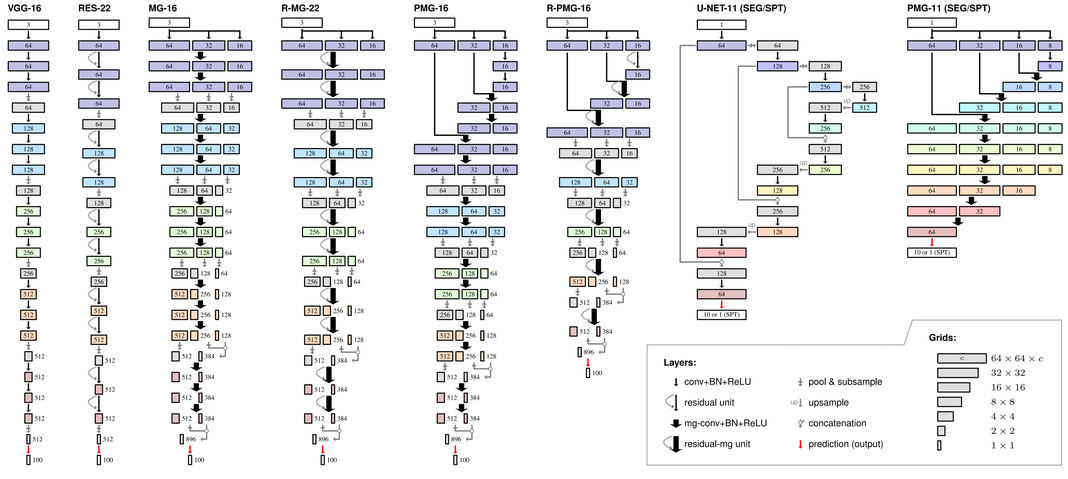


IMAGE CLASSIFICATION

Ablation Study (CIFAR):

- Multigrid capacity useful; performance improves over standard VGG nets
- Synergistic with residual connections
- Progressive versions:
  - save parameters
  - save FLOPs

Large Scale (ImageNet):

- R-MG-34 outperforms the deeper ResNet-50
- Multigrid *more efficient than going wider*; compare R-MG-34 with WRN-34
- R-PMG-30-SEG: unifying architecture for both segmentation & classification

CIFAR-100 Method	Params ($\times 10^6$)	FLOPs ($\times 10^6$)	Error (%)
VGG-6	3.96	67.88	32.15
VGG-11	9.46	228.31	30.26
VGG-16	14.96	388.75	30.47
VGG-21	20.46	549.18	31.89
MG-6	8.34	116.63	32.08
MG-11	20.46	391.88	28.39
MG-16	32.58	667.13	29.91
MG-21	44.69	942.38	30.03
PMG-9	8.46	186.21	30.61
PMG-16	20.60	468.09	28.11
PMG-30	32.74	749.98	29.89
RES-12	9.50	266.02	28.64
RES-22	20.49	586.93	28.05
RES-32	31.49	907.79	27.12
RES-42	42.48	1228.65	27.60
R-MG-12	20.56	457.20	27.84
R-MG-22	44.79	1007.70	26.79
R-MG-32	69.02	1558.20	25.29
R-MG-42	93.26	2108.71	26.32
R-PMG-16	20.60	468.09	27.35
R-PMG-30	44.88	1031.87	26.44
R-PMG-44	69.16	1595.64	26.68

ImageNet Method	Params ($\times 10^6$)	FLOPs ($\times 10^9$)	val, 10-crop Top-1	val, 10-crop Top-5	val, single-crop Top-1	val, single-crop Top-5
VGG-16 [7]	138.0	15.47	28.07	9.33	-	-
ResNet-34 C [2]	21.8	3.66	24.19	7.40	-	-
ResNet-50 [2]	25.6	4.46	22.85	6.71	-	-
WRN-34 (2.0) [6]	48.6	14.09	-	-	**24.50**	7.58
R-MG-34	32.9	5.76	**22.42**	**6.12**	**24.51**	**7.46**
R-PMG-30-SEG	31.9	2.77	23.60	6.89	26.50	8.63

SEMANTIC SEGMENTATION (SYNTHETIC)

Task: Separate superimposed digits into per-class channels

Dataset:

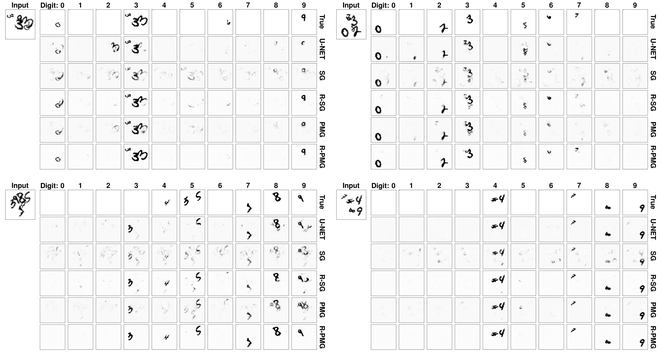
- Randomly distorted and overlapped MNIST digits
- 10K training, 1K test images; each 64×64 pixels

Results:

- Progressive multigrid nets perform dramatically better
- As in classification, synergistic with residual links

Method	Params ($\times 10^6$)	mIoU (%)	Error (%)
U-NET-11	3.79	58.46	22.14
U-MG-11	5.90	58.02	21.45
SG-11	0.23	32.21	22.33
R-SG-20	0.45	69.86	14.43
PMG-11	0.61	61.88	14.75
R-PMG-20	1.20	**81.91**	**8.89**

Performance on synthetic MNIST segmentation task.



SPATIAL TRANSFORMERS & LEARNED ATTENTION

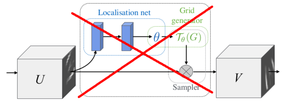
Task:

- Directly train CNNs to invert spatial transformations
- Output original instance in canonical pose

Dataset:

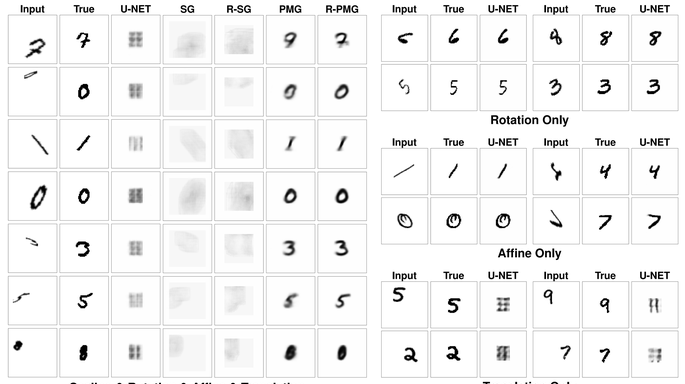
- Distorted MNIST digits: scaling, rotation, affine, & translation
- 60K training, 10K test images; single digit in 64×64 canvas

Results: Multigrid network structure is required to learn the task!



Spatial Transformer [3]

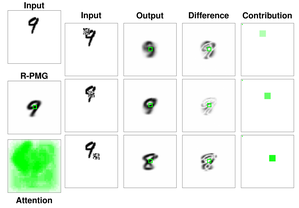
Idea: Replace with multigrid CNN & learn end-to-end.

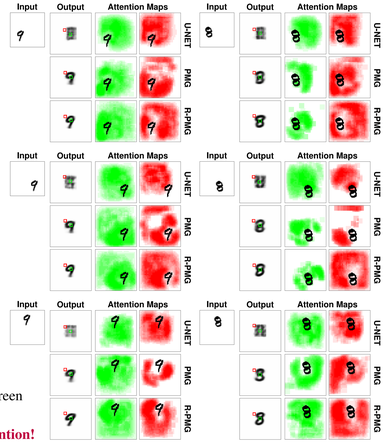


Method	Params ($\times 10^6$)	mIoU (%)	Error (%)
U-NET-11	3.79	12.20	37.55
SG-11	0.23	0	100.00
R-SG-20	0.45	0	82.98
PMG-11	0.61	**50.64**	29.20
R-PMG-20	1.20	**55.61**	25.59

Spatial transformer accuracy: only multigrid nets work!

INTROSPECTION



- Slide distractor (random patch) over image
- Record difference in output at specific location
- Generates attention map for that location (*e.g.* green square); shown is center-specific attention map

Only multigrid nets exhibit data-dependent attention!