Multigrid Neural Architectures

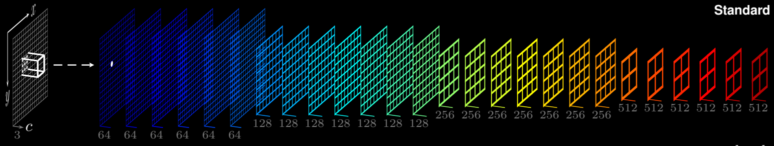
Tsung-Wei Ke
UC Berkeley / ICSI

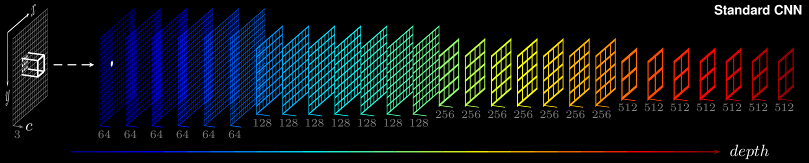
Michael Maire
TTI Chicago

Stella X. Yu
UC Berkeley / ICSI

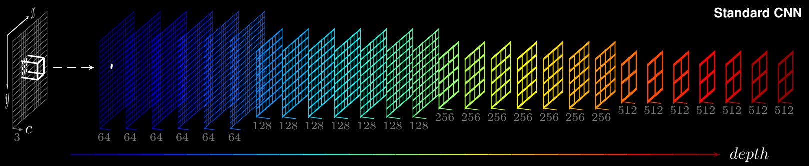
arXiv:1611.07661 & CVPR 2017

Standard CNN

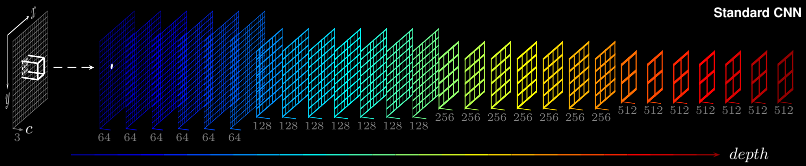




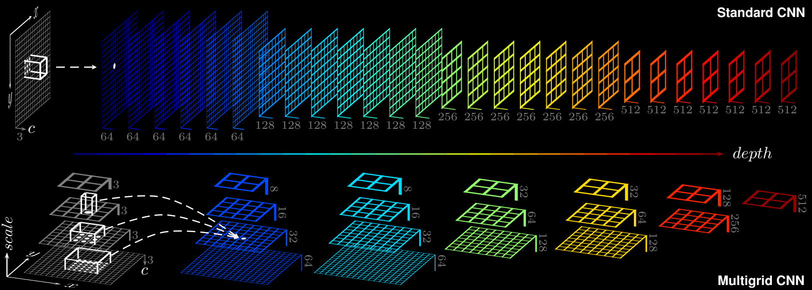
- Conflates abstraction & scale



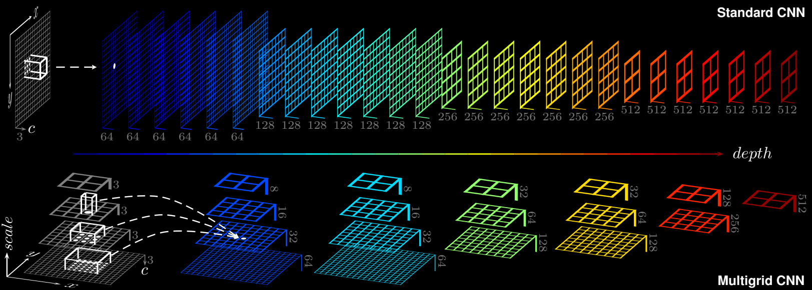
- Conflates abstraction & scale
- Fine-to-coarse



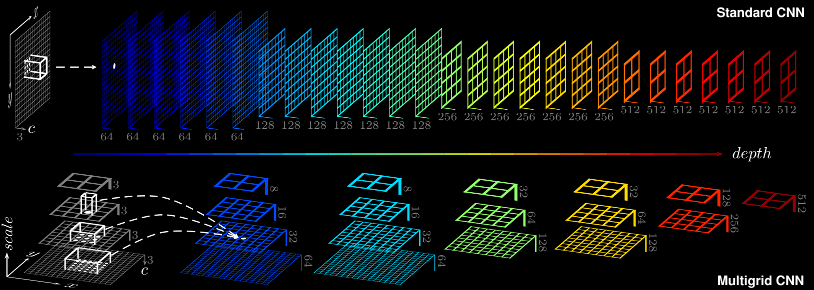
- Conflates abstraction & scale
- Fine-to-coarse
- Slow receptive field growth



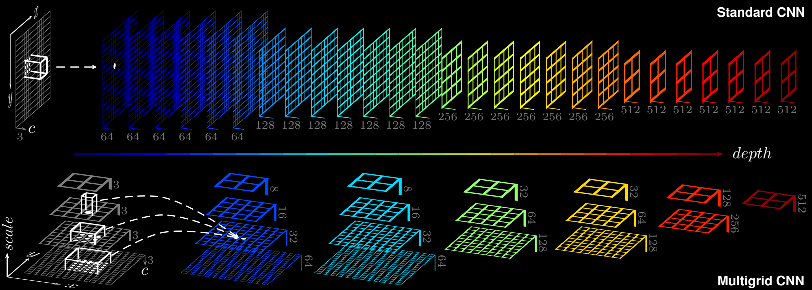
- Conflates abstraction & scale
- Fine-to-coarse
- Slow receptive field growth



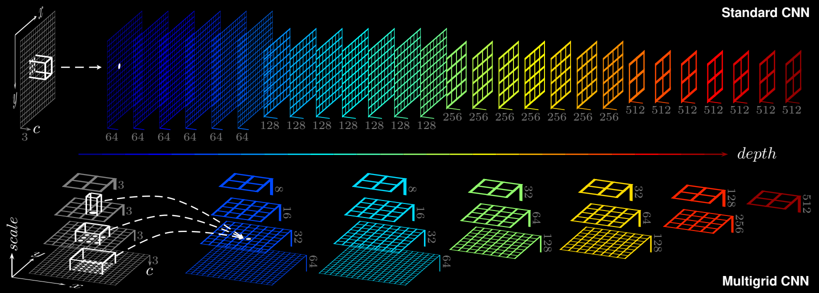
- Conflates abstraction & scale → disentangles them
- Fine-to-coarse
- Slow receptive field growth



- Conflates abstraction & scale → disentangles them
- Fine-to-coarse → coarse & fine (optionally progressive)
- Slow receptive field growth



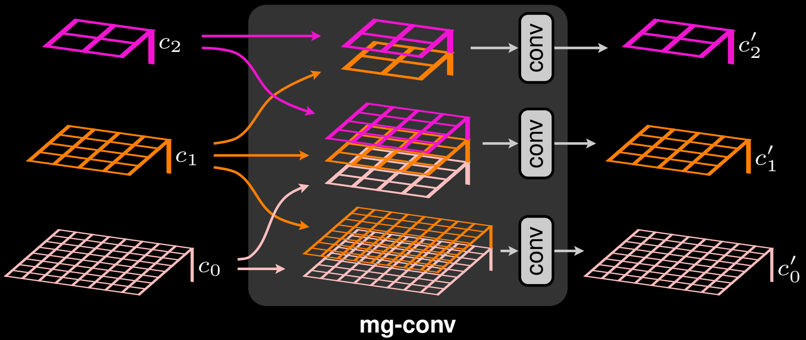
- Conflates abstraction & scale → disentangles them
- Fine-to-coarse → coarse & fine (optionally progressive)
- Slow receptive field growth → efficient communication



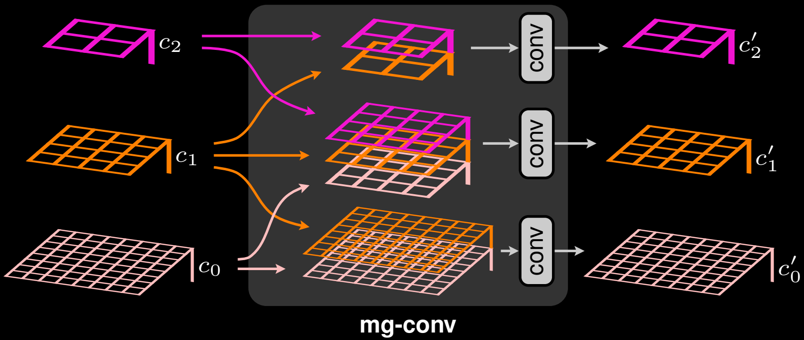
- Conflates abstraction & scale → disentangles them
- Fine-to-coarse → coarse & fine (optionally progressive)
- Slow receptive field growth → efficient communication

Multigrid destroys the notion of receptive field

Multigrid Convolution

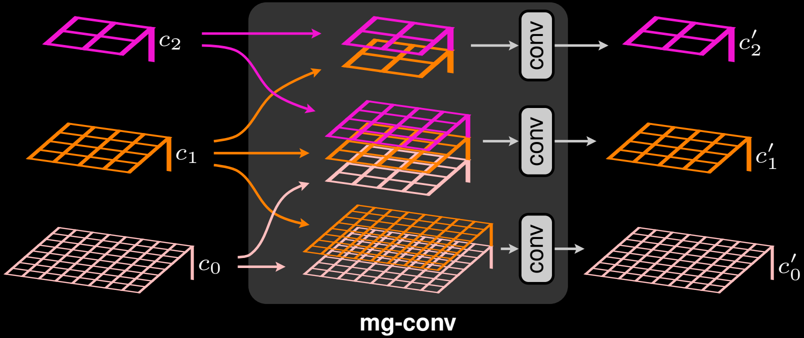


Multigrid Convolution



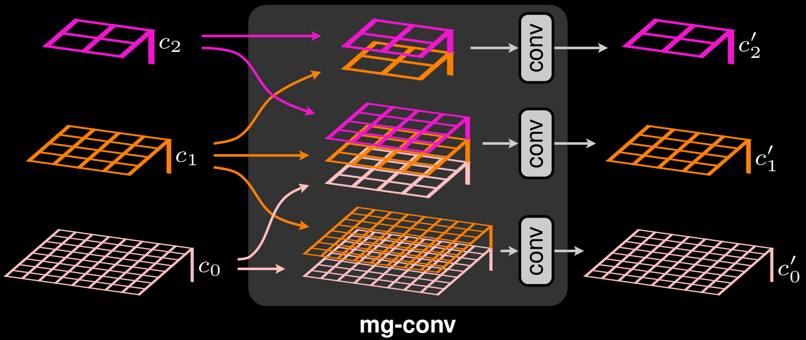
- Upsampling: nearest neighbor

Multigrid Convolution



- Upsampling: nearest neighbor
- Downsampling: max-pooling

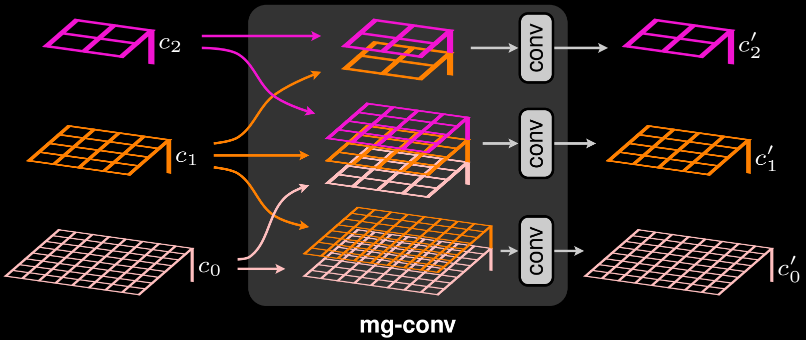
Multigrid Convolution



- Upsampling: nearest neighbor
- Downsampling: max-pooling

Rethink pooling: communication, not summarization

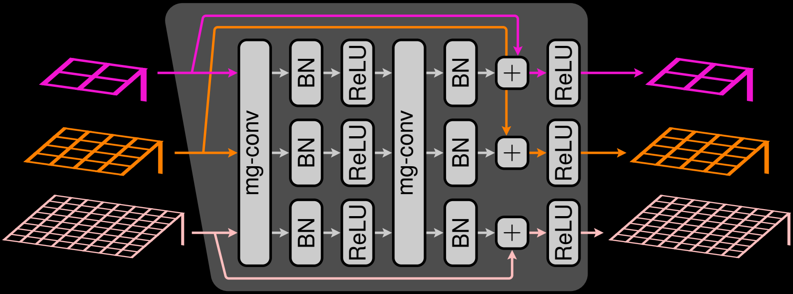
Multigrid Convolution



- Upsampling: nearest neighbor
- Downsampling: max-pooling

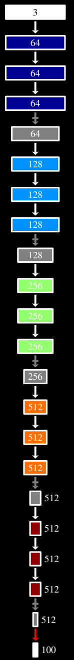
CNN evolves representation on pyramid

Multigrid Residual Unit

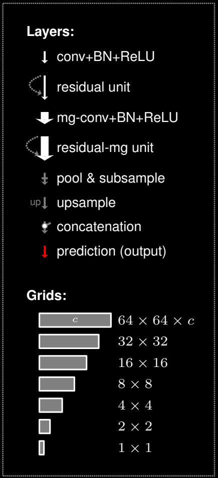


res-mg-unit

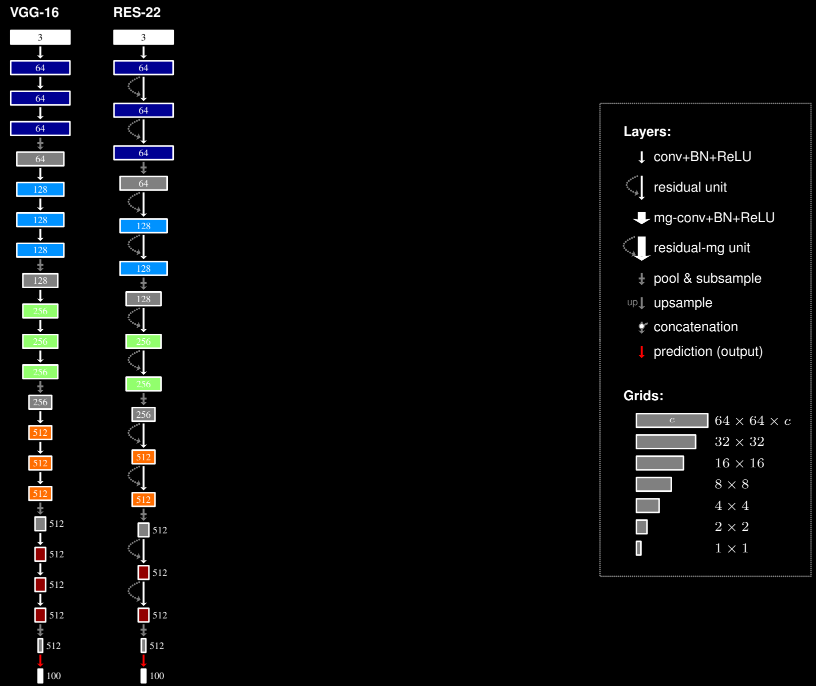
VGG-16

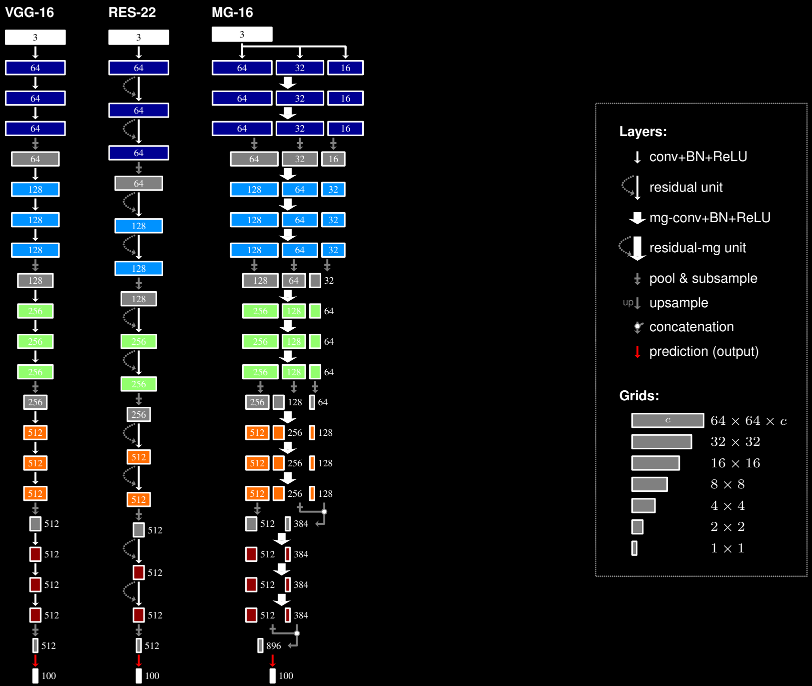


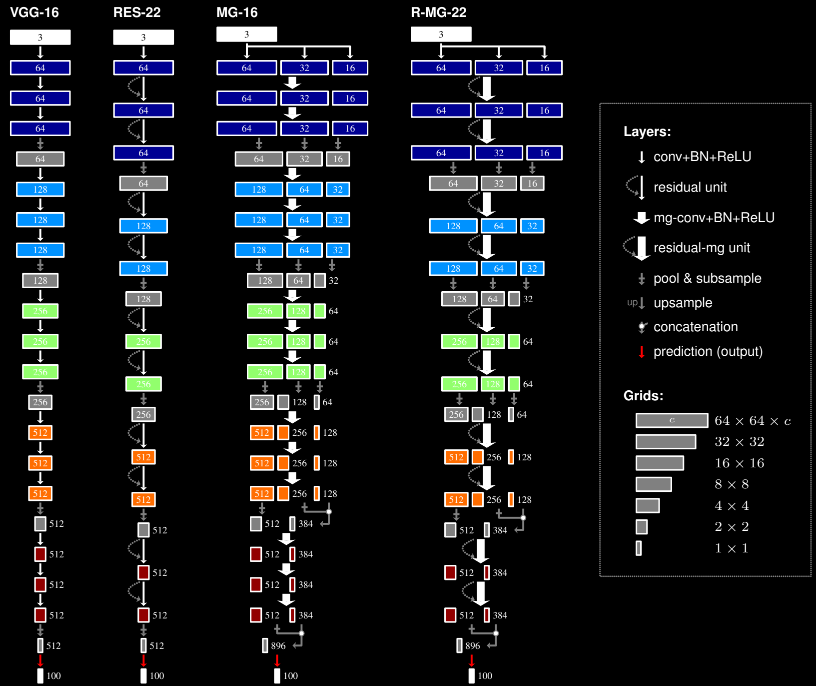
3
64
64
64
64
128
128
128
128
256
256
256
256
512
512
512
512
512
512
512
512
100



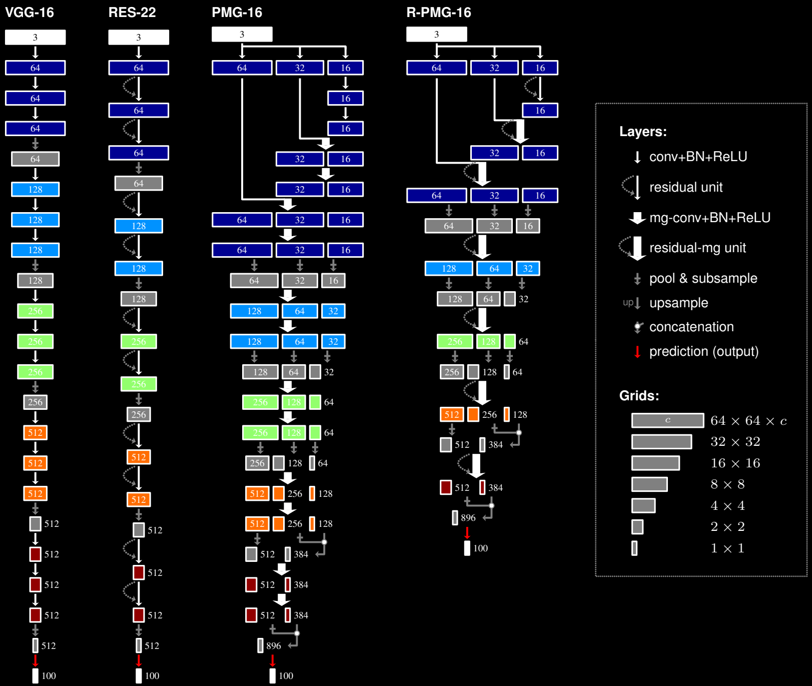
Layers:
conv+BN+ReLU
residual unit
mg-conv+BN+ReLU
residual-mg unit
pool & subsample
up
upsample
concatenation
prediction (output)
Grids:
c
64 × 64 × c
32 × 32
16 × 16
8 × 8
4 × 4
2 × 2
1 × 1


VGG-16
RES-22
3
64
128
256
512
100
Layers:
conv+BN+ReLU
residual unit
mg-conv+BN+ReLU
residual-mg unit
pool & subsample
up
upsample
concatenation
prediction (output)
Grids:
c
64 × 64 × c
32 × 32
16 × 16
8 × 8
4 × 4
2 × 2
1 × 1


VGG-16
RES-22
MG-16
Layers:
conv+BN+ReLU
residual unit
mg-conv+BN+ReLU
residual-mg unit
pool & subsample
up
upsample
concatenation
prediction (output)
Grids:
$64 \times 64 \times c$
32×32
16×16
8×8
4×4
2×2
1×1



VGG-16
RES-22
MG-16
R-MG-22
Layers:
conv+BN+ReLU
residual unit
mg-conv+BN+ReLU
residual-mg unit
pool & subsample
up
upsample
concatenation
prediction (output)
Grids:
c
64 × 64 × c
32 × 32
16 × 16
8 × 8
4 × 4
2 × 2
1 × 1


VGG-16
RES-22
PMG-16
R-PMG-16
Layers:
conv+BN+ReLU
residual unit
mg-conv+BN+ReLU
residual-mg unit
pool & subsample
up
upsample
concatenation
prediction (output)
Grids:
c
$64 \times 64 \times c$
32×32
16×16
8×8
4×4
2×2
1×1

Results: CIFAR 100 Classification

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

Results: CIFAR 100 Classification

- Multigrid capacity useful, improves performance

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

Results: CIFAR 100 Classification

- Multigrid capacity useful, improves performance
- Synergistic with residual connections

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

Results: CIFAR 100 Classification

- Multigrid capacity useful, improves performance
- Synergistic with residual connections
- Progressive versions:
  - save params
  - save FLOPs

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

Results: ImageNet Classification

Method	Params ($\times 10^6$)	FLOPs ($\times 10^9$)	val, 10-crop Top-1	val, 10-crop Top-5
VGG-16 (Simonyan and Zisserman, 2015)	138.0	15.47	28.07	9.33
ResNet-34 C (He et al., 2016)	21.8	3.66	24.19	7.40
ResNet-50 (He et al., 2016)	25.6	4.46	22.85	6.71
WRN-34 (2.0) (Zagoruyko and Komodakis, 2016)	48.6	14.09	-	-
R-MG-34	32.9	5.76	**22.42**	**6.12**
R-PMG-30-SEG	31.9	2.77	23.60	6.89

ImageNet validation set error, 10-crop

R-MG-34 outperforms ResNet-50

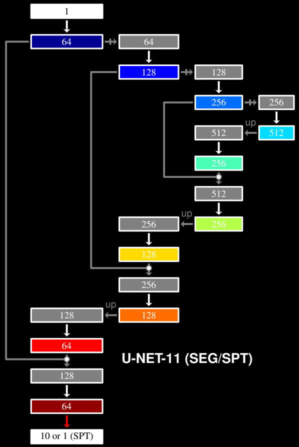
Results: ImageNet Classification

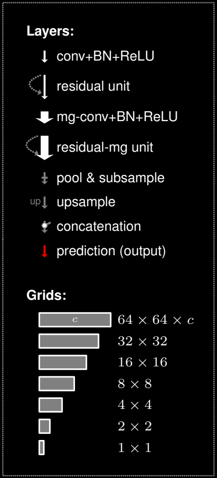
Method	Params ($\times 10^6$)	FLOPs ($\times 10^9$)	val, 1-crop Top-1	Top-5
VGG-16 (Simonyan and Zisserman, 2015)	138.0	15.47	-	-
ResNet-34 C (He et al., 2016)	21.8	3.66	-	-
ResNet-50 (He et al., 2016)	25.6	4.46	-	-
WRN-34 (2.0) (Zagoruyko and Komodakis, 2016)	48.6	14.09	**24.50**	7.58
R-MG-34	32.9	5.76	**24.51**	**7.46**
R-PMG-30-SEG	31.9	2.77	26.50	8.63

ImageNet validation set error, single-crop

R-MG-34 matches WRN-34, using fewer params and FLOPs

Networks for Semantic Segmentation

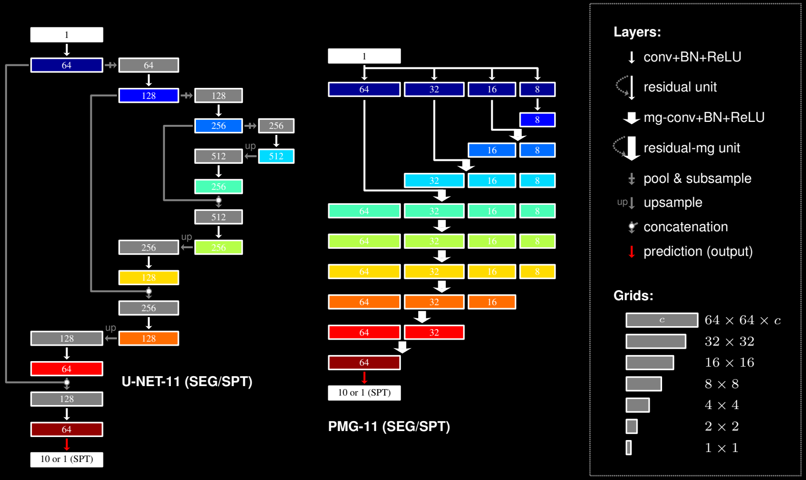




U-NET: Ronneberger et al. (2015)
similar: autoencoders; Badrinarayanan et al. (2015);

Hu and Ramanan (2016); Newell et al. (2016)

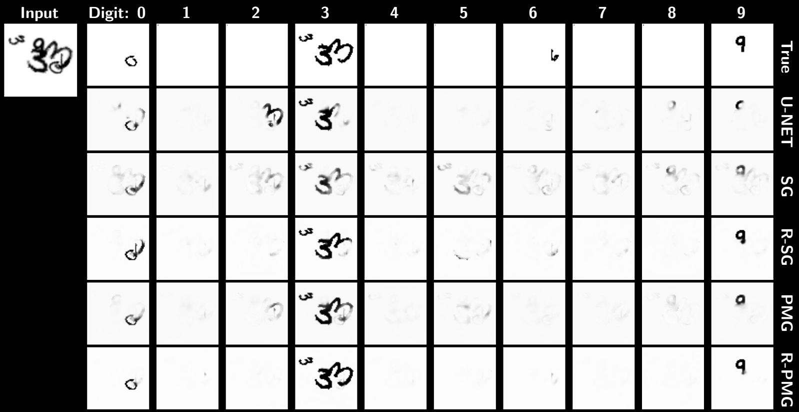
Networks for Semantic Segmentation



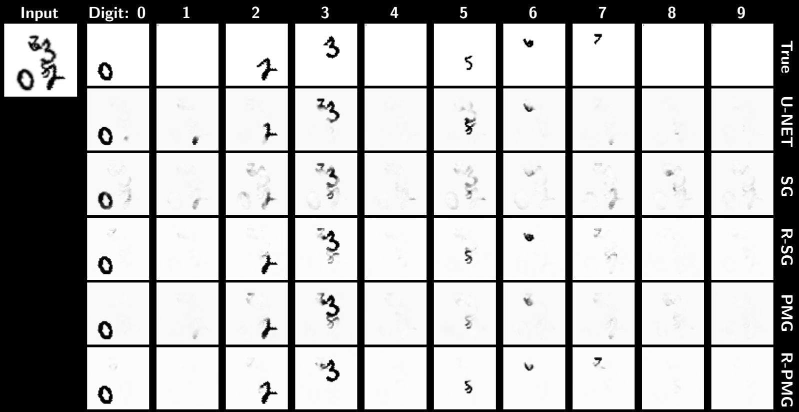
U-NET: Ronneberger et al. (2015)
similar: autoencoders; Badrinarayanan et al. (2015);

Hu and Ramanan (2016); Newell et al. (2016)

MNIST Synthetic Semantic Segmentation



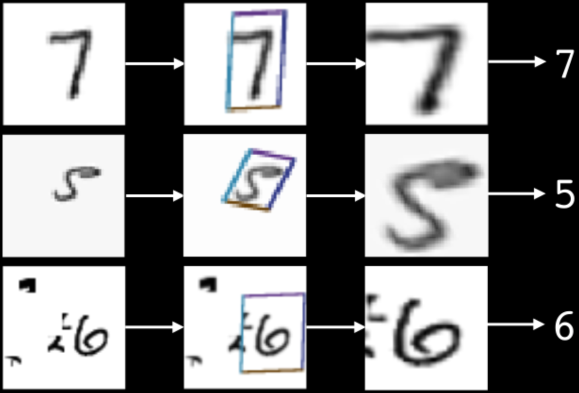
MNIST Synthetic Semantic Segmentation



MNIST Synthetic Semantic Segmentation

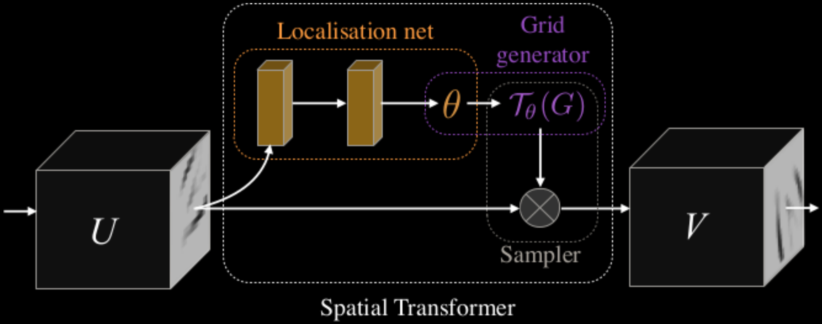
Method	Params ($\times 10^6$)	IoU (%)	Error (%)
U-NET-11	3.79	58.46	22.14
U-MG-11	5.90	58.02	21.45
SG-11	0.23	32.21	22.33
R-SG-20	0.45	69.86	14.43
PMG-11	0.61	61.88	14.75
R-PMG-20	1.20	81.91	8.89

Spatial Transformations



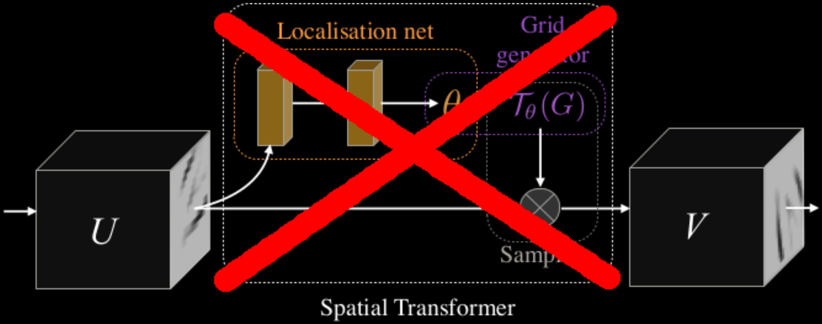
Jaderberg et al. (2015)

Spatial Transformations



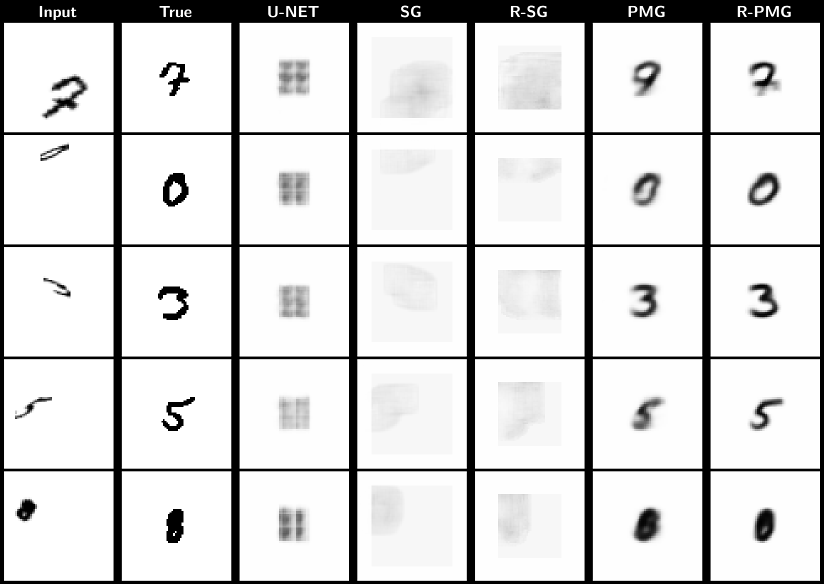
Jaderberg et al. (2015)

Spatial Transformations



Replace with Multigrid CNN

MNIST Spatial Transformation Task



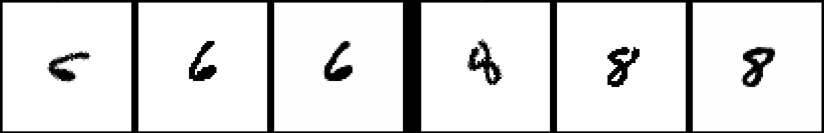
MNIST Spatial Transformation Task

Method	Params ($\times 10^6$)	IoU (%)	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

Multigrid necessary!

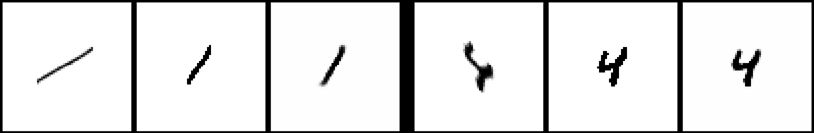
MNIST Spatial Transformation Task

Input | True | U-NET | Input | True | U-NET



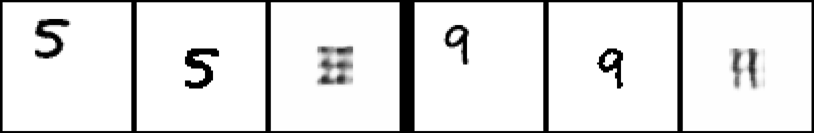
Rotation Only

Input | True | U-NET | Input | True | U-NET



Affine Only

Input | True | U-NET | Input | True | U-NET

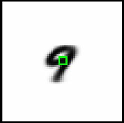


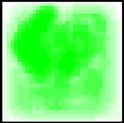
Translation Only

Multigrid Enables Learned Attention

Input

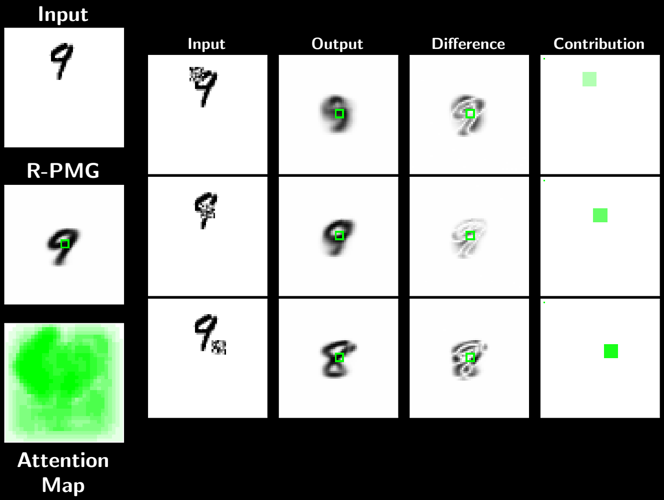
R-PMG



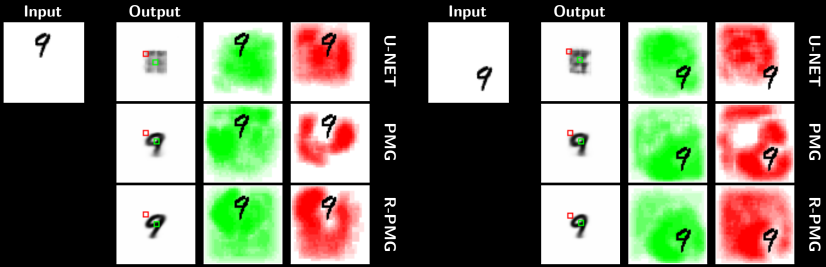


Attention Map

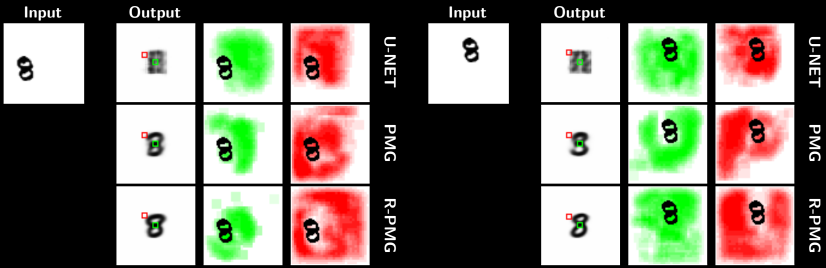
Multigrid Enables Learned Attention

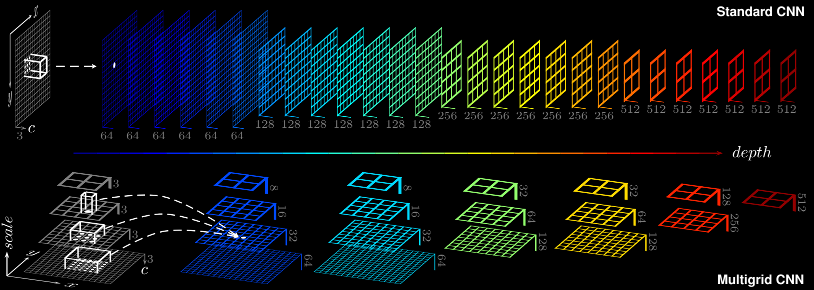


Multigrid Enables Learned Attention



Multigrid Enables Learned Attention





- Representation: evolve scale-space features
- Computation: progressive coarse-to-fine
- Communication: enable dynamic routing & attention