

Redes Neurais e Aprendizagem Profunda

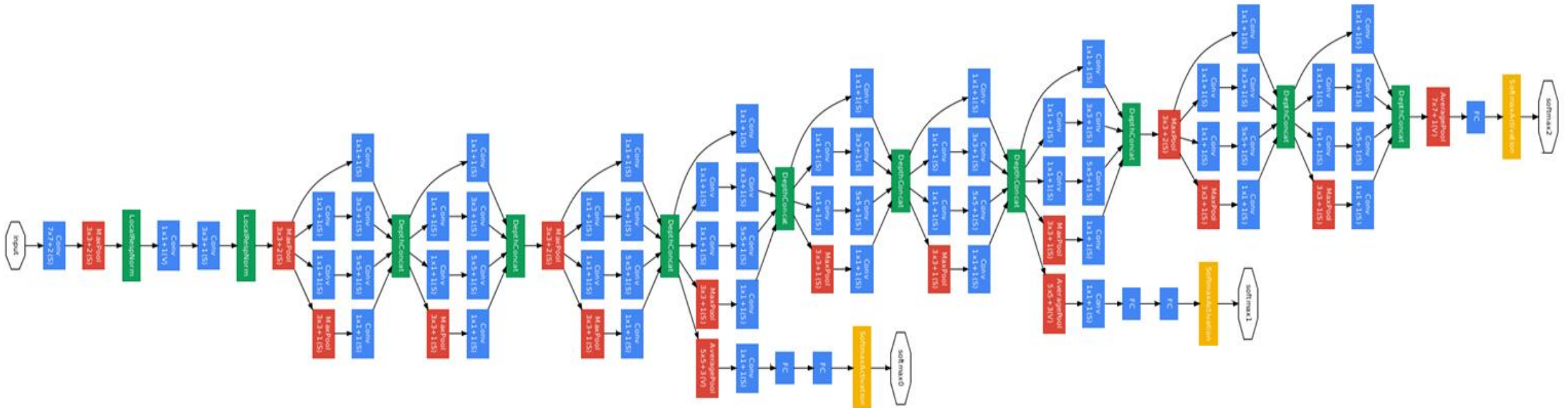
REDES NEURAIS CONVOLUCIONAIS GOOGLNET / RESNET / ANÁLISE

Zenilton K. G. Patrocínio Jr
zenilton@pucminas.br

ConvNets – GoogLeNet

[Szegedy et al., 2014]

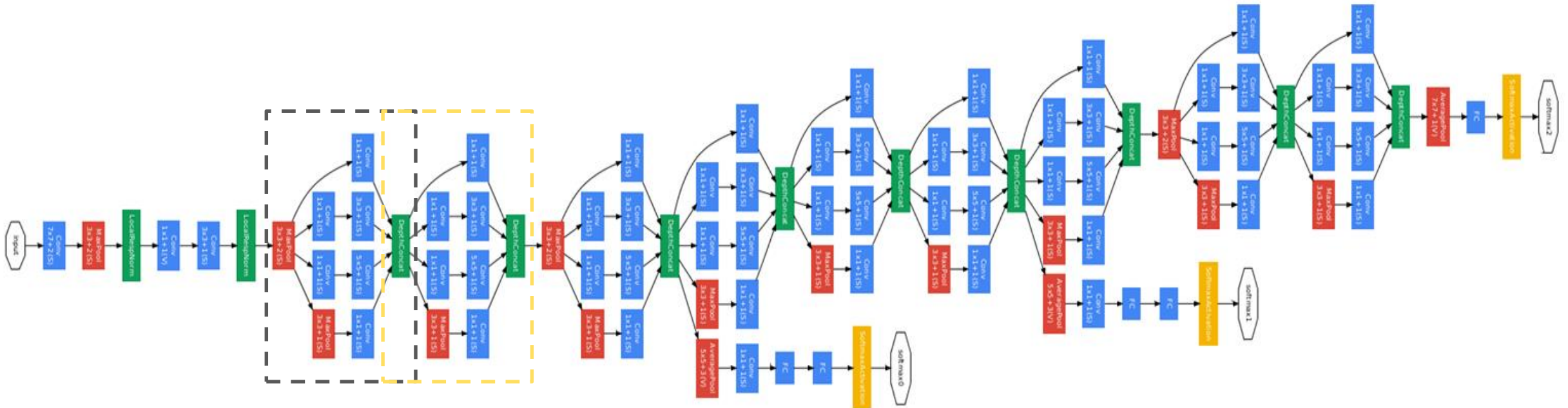
Uso de vários filtros em paralelo



ConvNets – GoogLeNet

[Szegedy et al., 2014]

Uso de vários filtros em paralelo

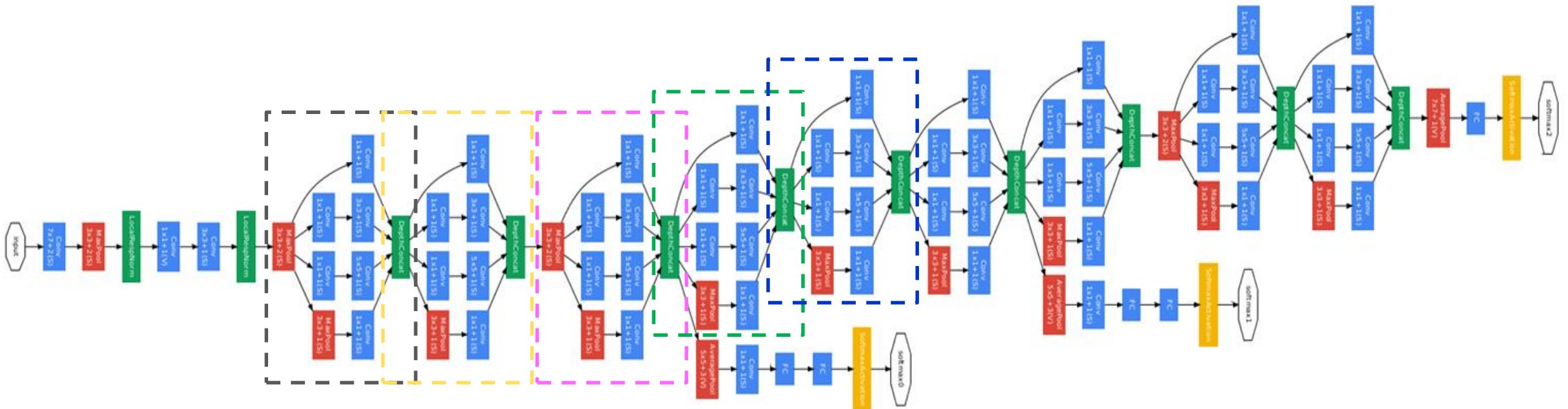


Módulo *Inception*

ConvNets – GoogLeNet

[Szegedy et al., 2014]

Uso de vários filtros em paralelo

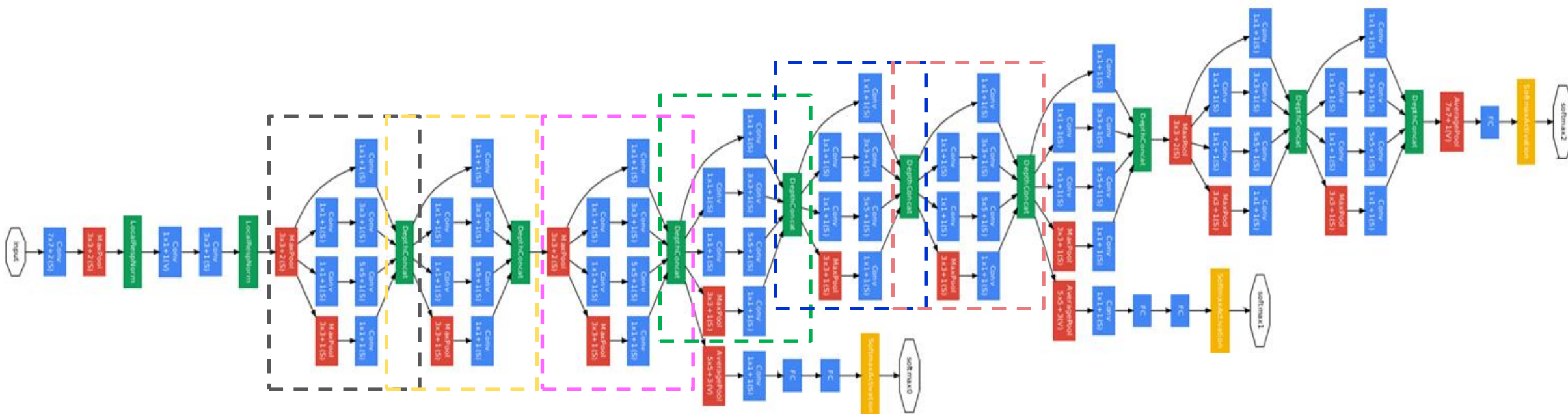


Módulo *Inception*

ConvNets – GoogLeNet

[Szegedy et al., 2014]

Uso de vários filtros em paralelo

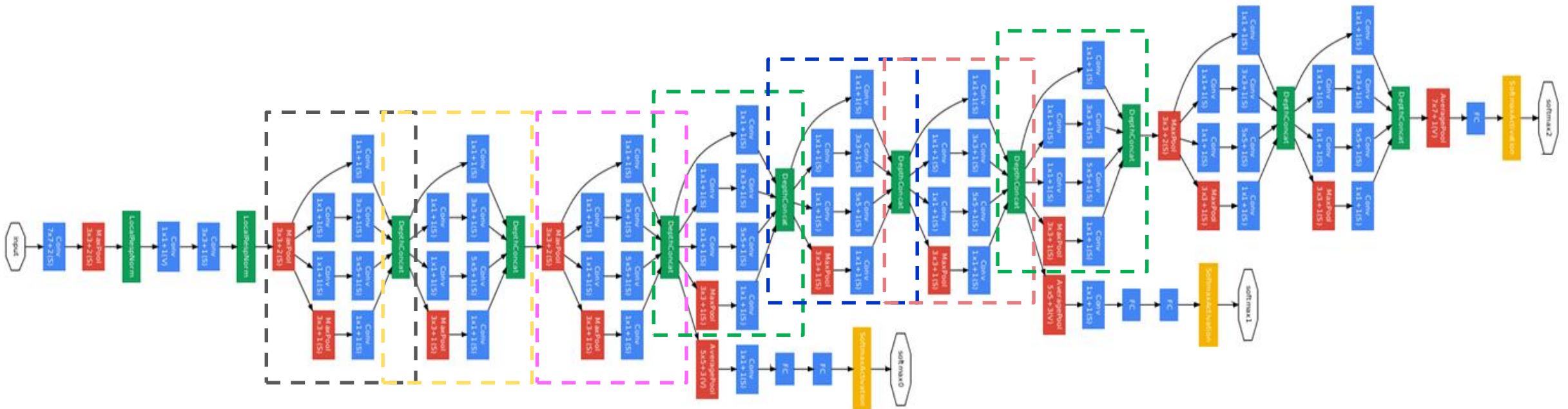


Módulo *Inception*

ConvNets – GoogLeNet

[Szegedy et al., 2014]

Uso de vários filtros em paralelo

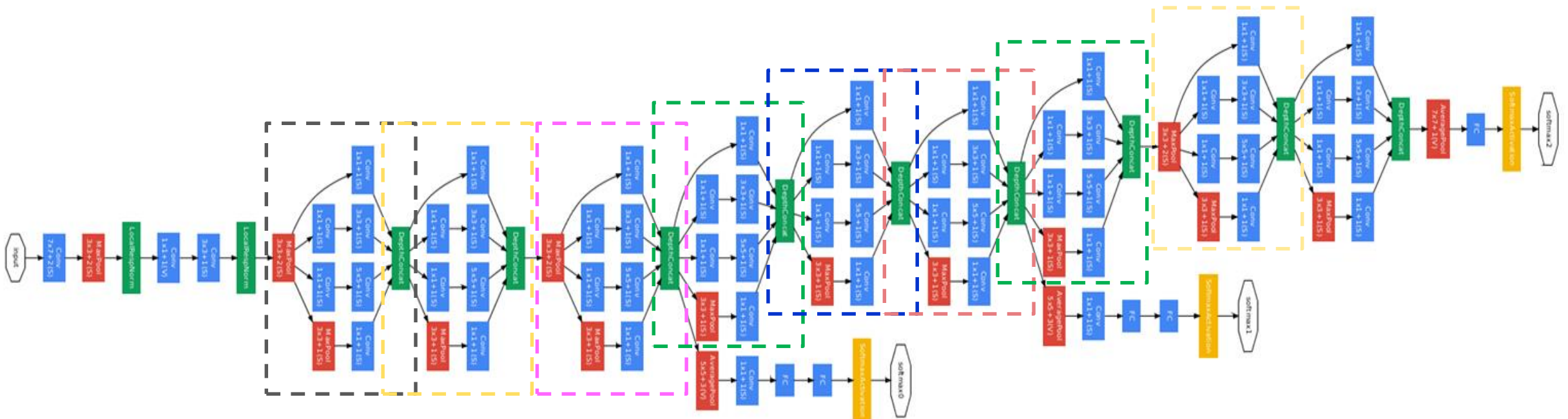


Módulo *Inception*

ConvNets – GoogLeNet

[Szegedy et al., 2014]

Uso de vários filtros em paralelo

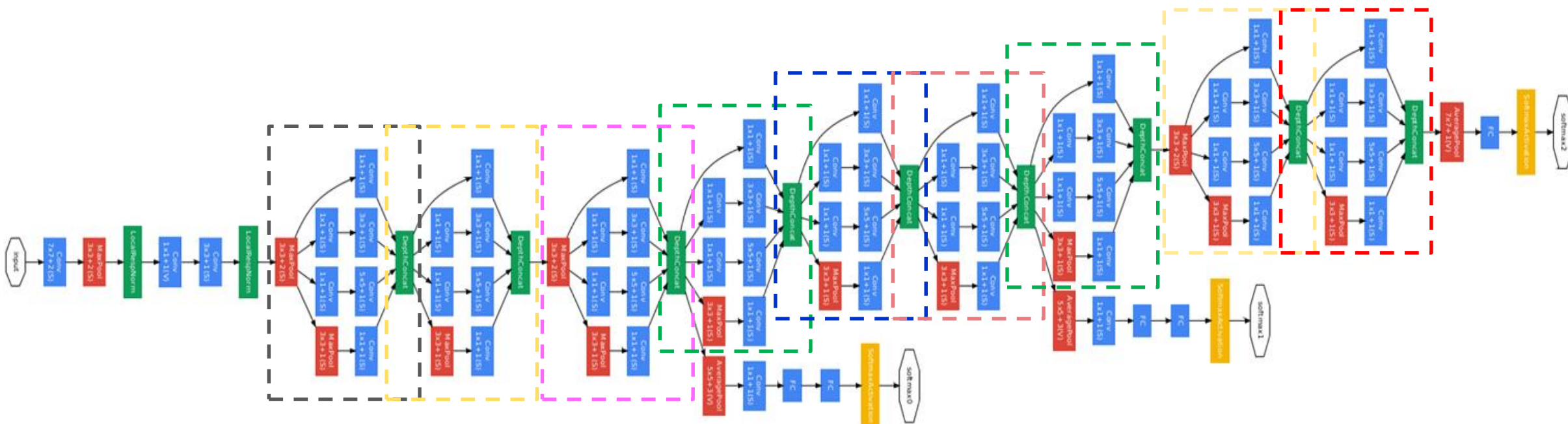


Módulo *Inception*

ConvNets – GoogLeNet

[Szegedy et al., 2014]

Uso de vários filtros em paralelo

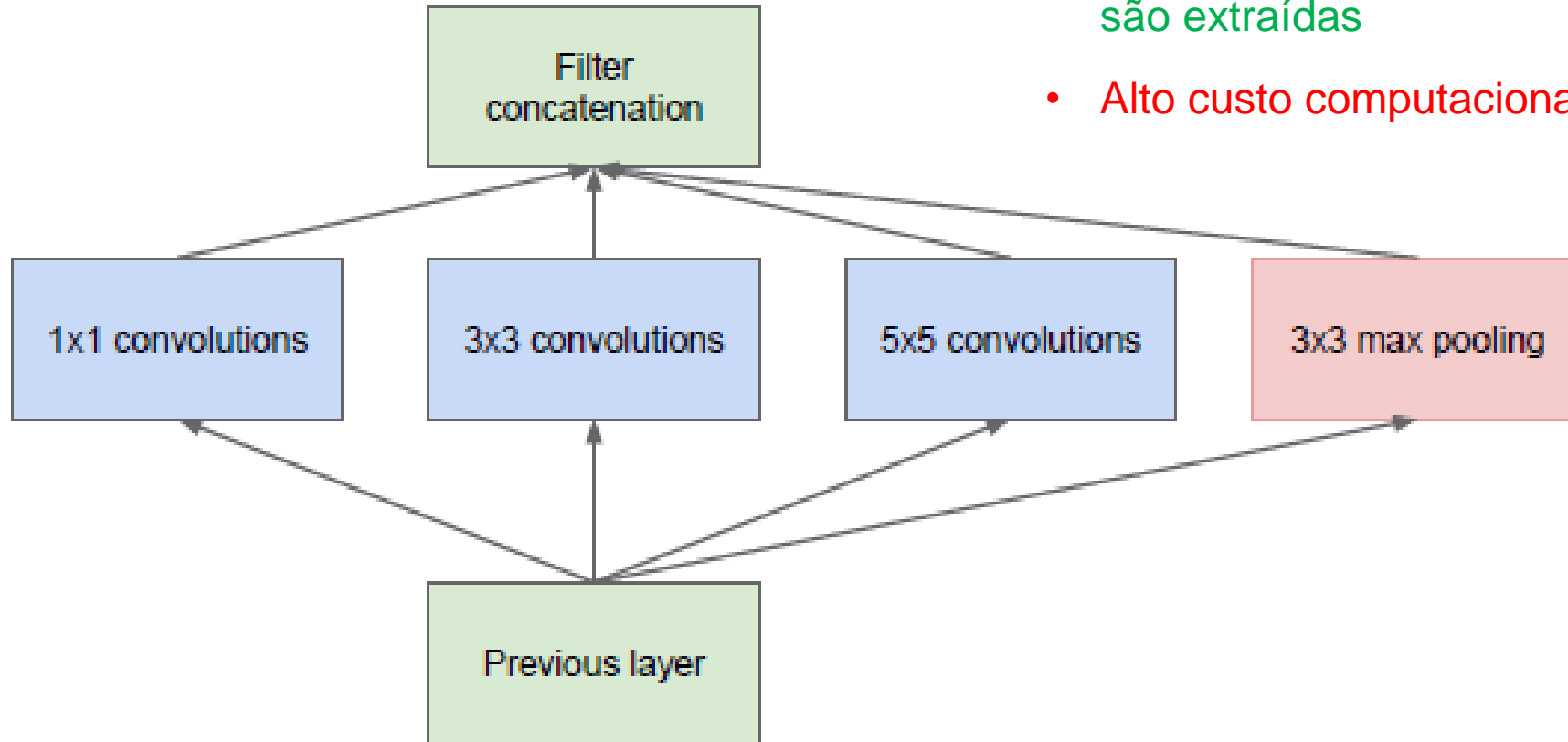


Módulo *Inception*

ConvNets – GoogLeNet

[Szegedy et al., 2014]

Módulo *Inception* – Versão básica

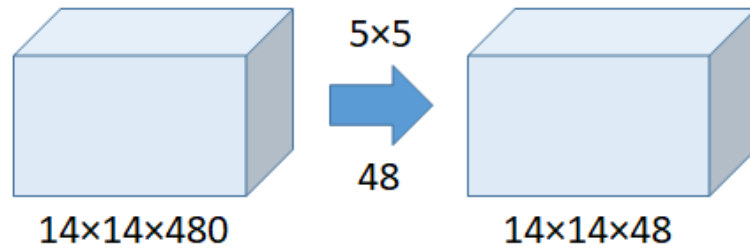


- Diferentes tipos de características são extraídas
- Alto custo computacional

ConvNets – GoogLeNet

[Szegedy et al., 2014]

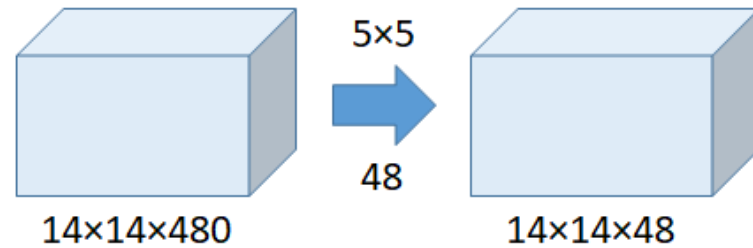
Módulo *Inception* – Versão básica



ConvNets – GoogLeNet

[Szegedy et al., 2014]

Módulo *Inception* – Versão básica

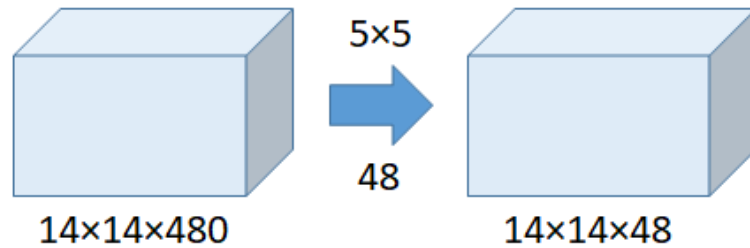


$$\begin{aligned}\text{Número de operações} &= (14 \times 14 \times 48) \times (5 \times 5 \times 480) = \\ &= 112,9\text{M}\end{aligned}$$

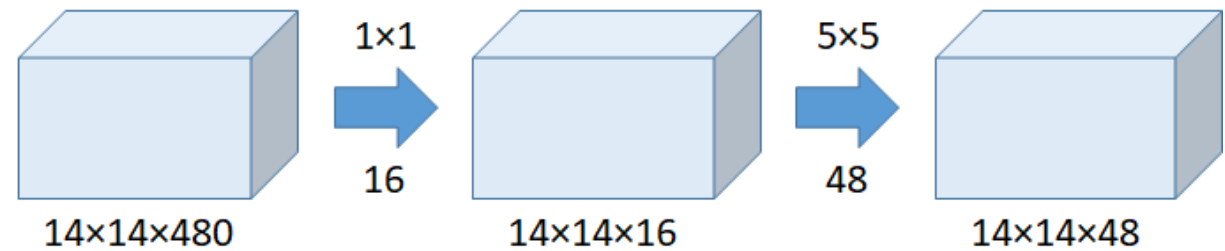
ConvNets – GoogLeNet

[Szegedy et al., 2014]

Módulo *Inception* – Versão básica



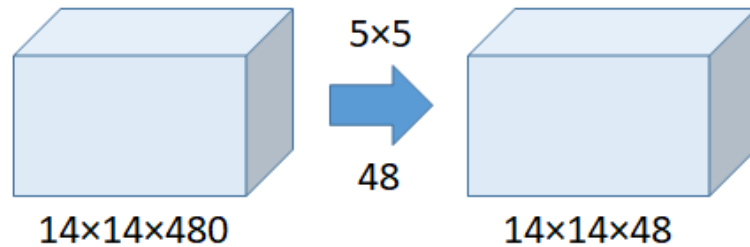
Número de operações = $(14 \times 14 \times 48) \times (5 \times 5 \times 480) =$
 $= 112,9\text{M}$



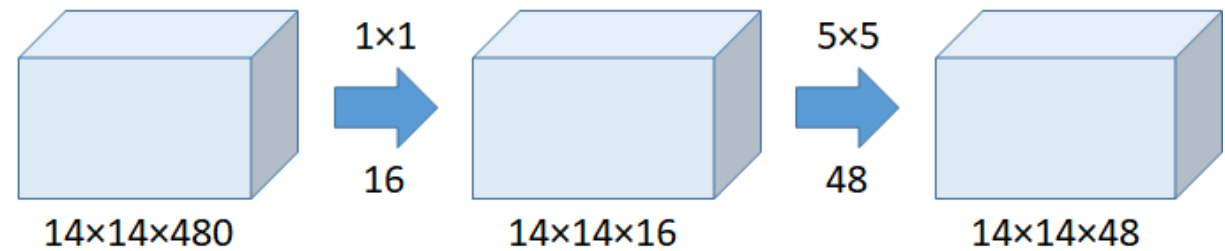
ConvNets – GoogLeNet

[Szegedy et al., 2014]

Módulo *Inception* – Versão básica



Número de operações = $(14 \times 14 \times 48) \times (5 \times 5 \times 480) =$
 $= 112,9\text{M}$

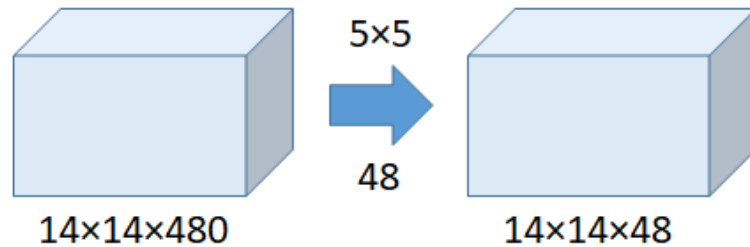


Núm. operações para $1 \times 1 = (14 \times 14 \times 16) \times (1 \times 1 \times 480) = 1,5\text{M}$

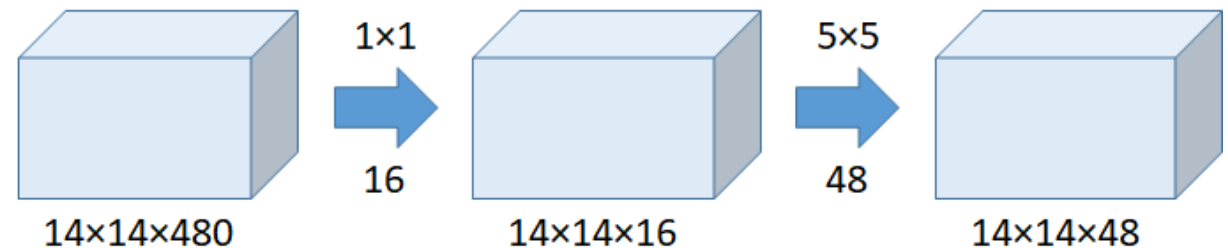
ConvNets – GoogLeNet

[Szegedy et al., 2014]

Módulo *Inception* – Versão básica



Número de operações = $(14 \times 14 \times 48) \times (5 \times 5 \times 480) =$
 $= 112,9\text{M}$



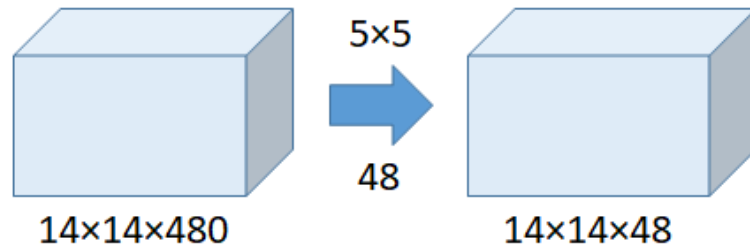
Núm. operações para $1 \times 1 = (14 \times 14 \times 16) \times (1 \times 1 \times 480) = 1,5\text{M}$

Núm. operações para $5 \times 5 = (14 \times 14 \times 48) \times (5 \times 5 \times 16) = 3,8\text{M}$

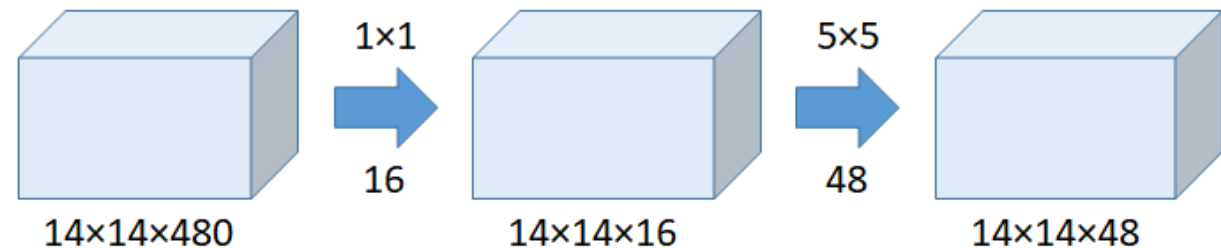
ConvNets – GoogLeNet

[Szegedy et al., 2014]

Módulo *Inception* – Versão básica



Número de operações = $(14 \times 14 \times 48) \times (5 \times 5 \times 480) =$
 $= \mathbf{112,9M}$



Núm. operações para $1 \times 1 = (14 \times 14 \times 16) \times (1 \times 1 \times 480) = 1,5M$

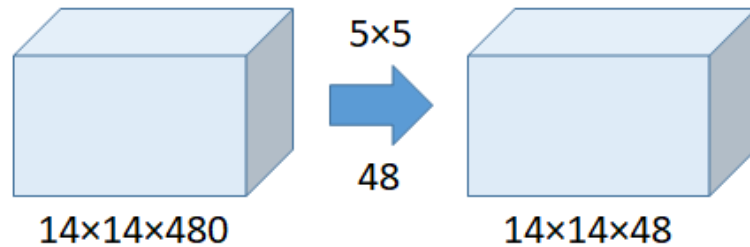
Núm. operações para $5 \times 5 = (14 \times 14 \times 48) \times (5 \times 5 \times 16) = 3,8M$

Núm. total de operações = $1,5M + 3,8M = \mathbf{5,3M}$

ConvNets – GoogLeNet

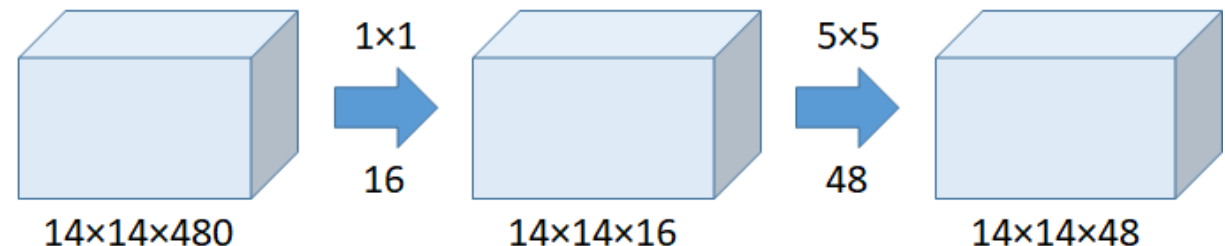
[Szegedy et al., 2014]

Módulo *Inception* – Versão básica



- CONV 1x1 redução de dimensionalidade
- Redução do custo computacional
- Aumento do número de parâmetros

Número de operações = $(14 \times 14 \times 48) \times (5 \times 5 \times 480) =$
= 112,9M



Núm. operações para 1×1 = $(14 \times 14 \times 16) \times (1 \times 1 \times 480) = 1,5\text{M}$

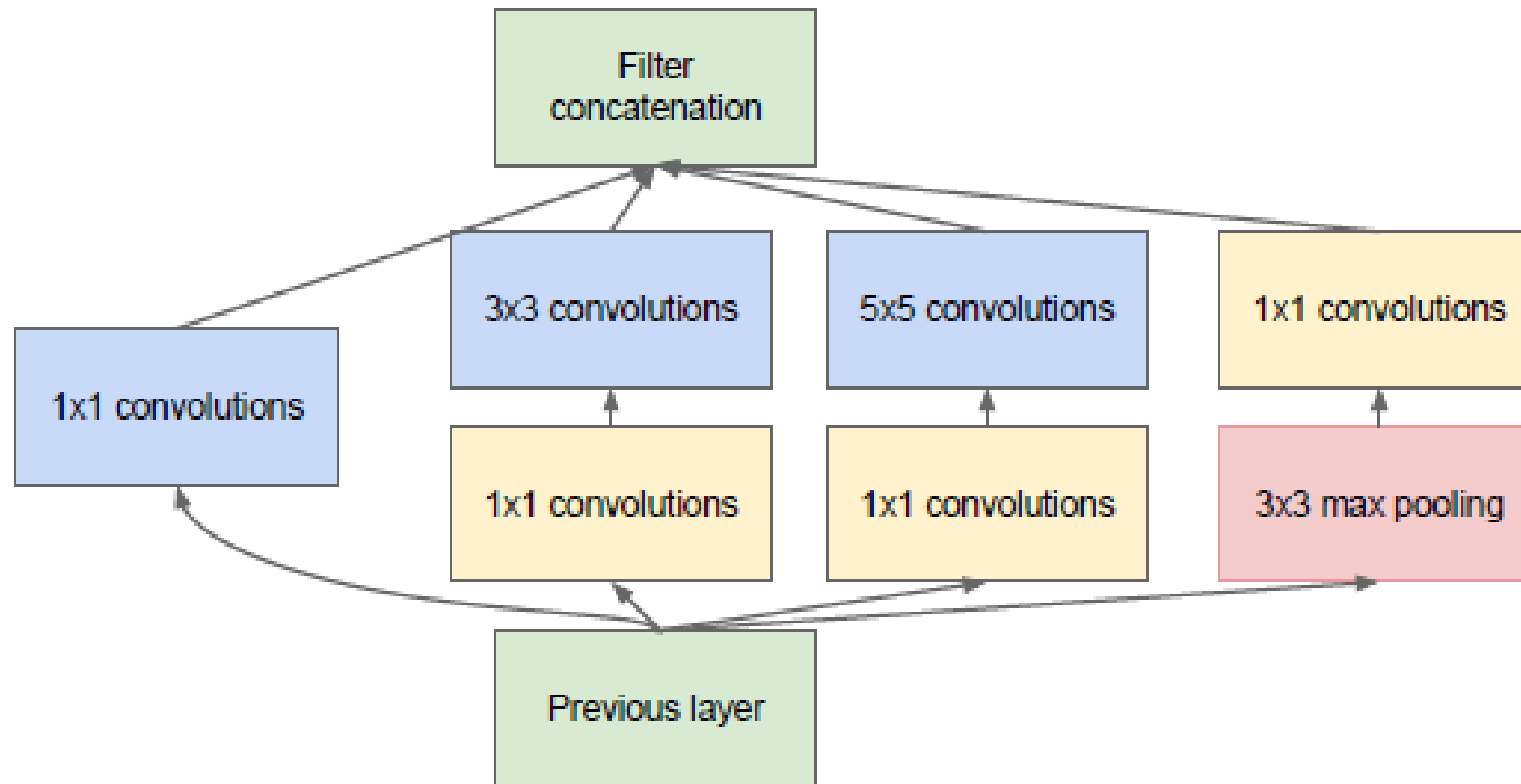
Núm. operações para 5×5 = $(14 \times 14 \times 48) \times (5 \times 5 \times 16) = 3,8\text{M}$

Núm. total de operações = $1,5\text{M} + 3,8\text{M} =$ **5,3M**

ConvNets – GoogLeNet

[Szegedy et al., 2014]

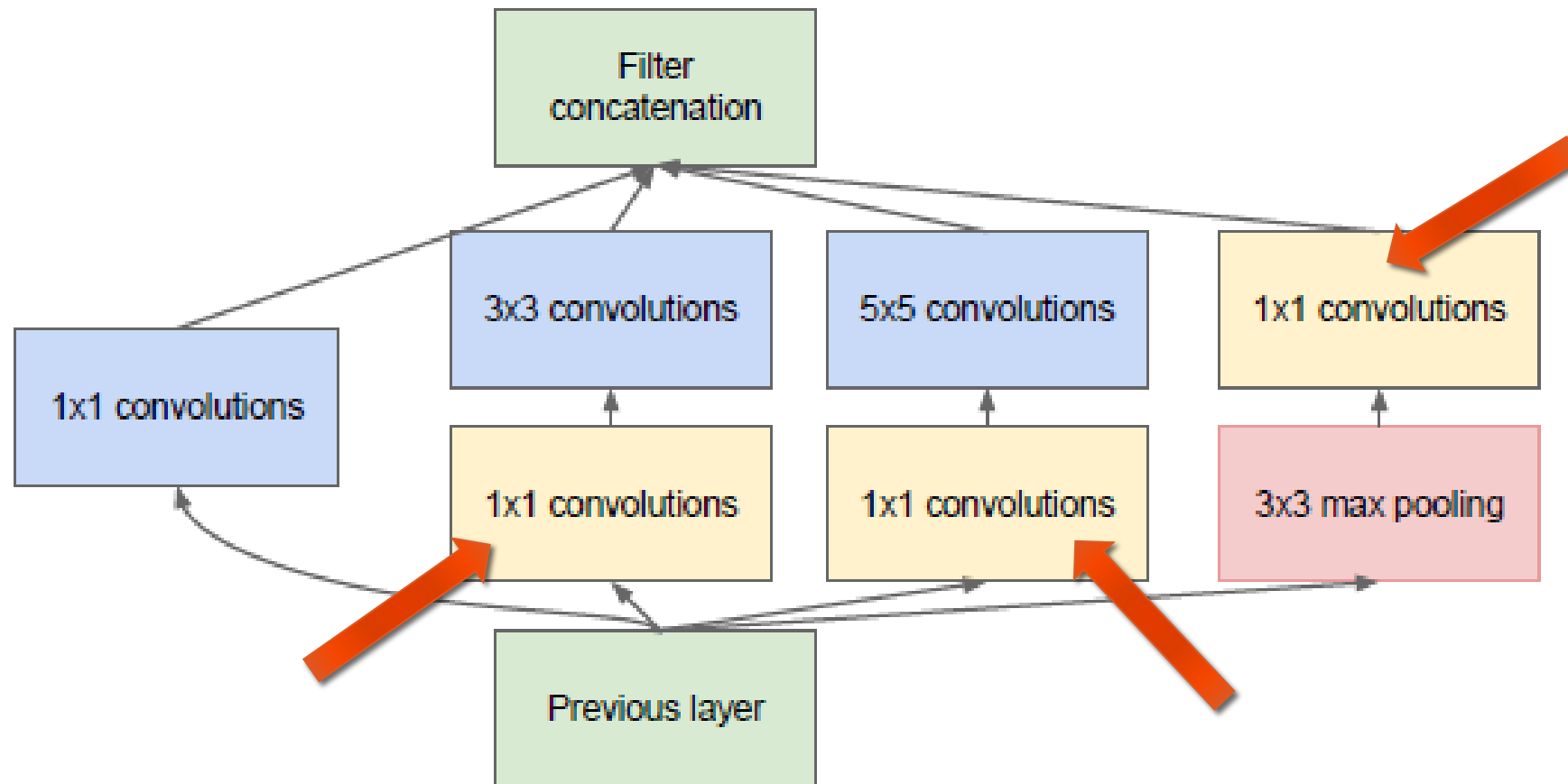
Módulo *Inception* – Versão completa



ConvNets – GoogLeNet

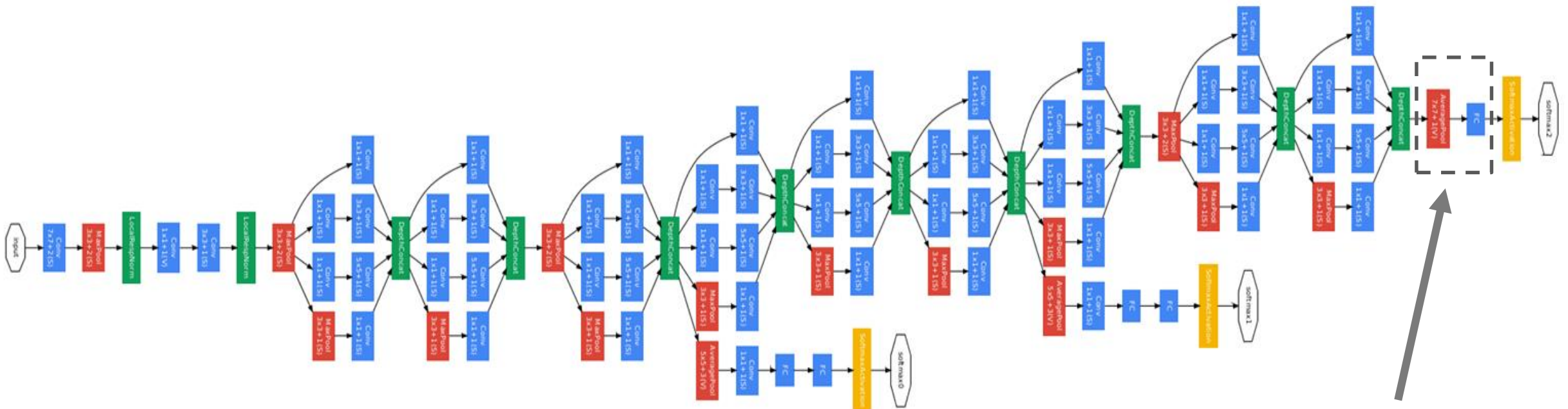
[Szegedy et al., 2014]

Módulo *Inception* – Versão completa



ConvNets – GoogLeNet

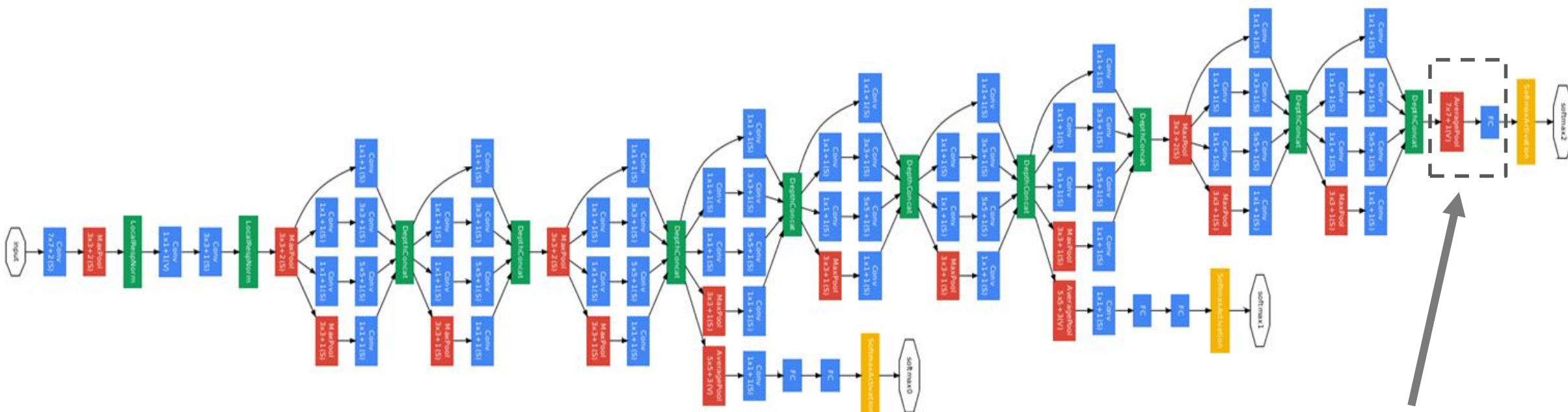
[Szegedy et al., 2014]



Agrupamento
pela Média
(average pooling)

ConvNets – GoogLeNet

[Szegedy et al., 2014]

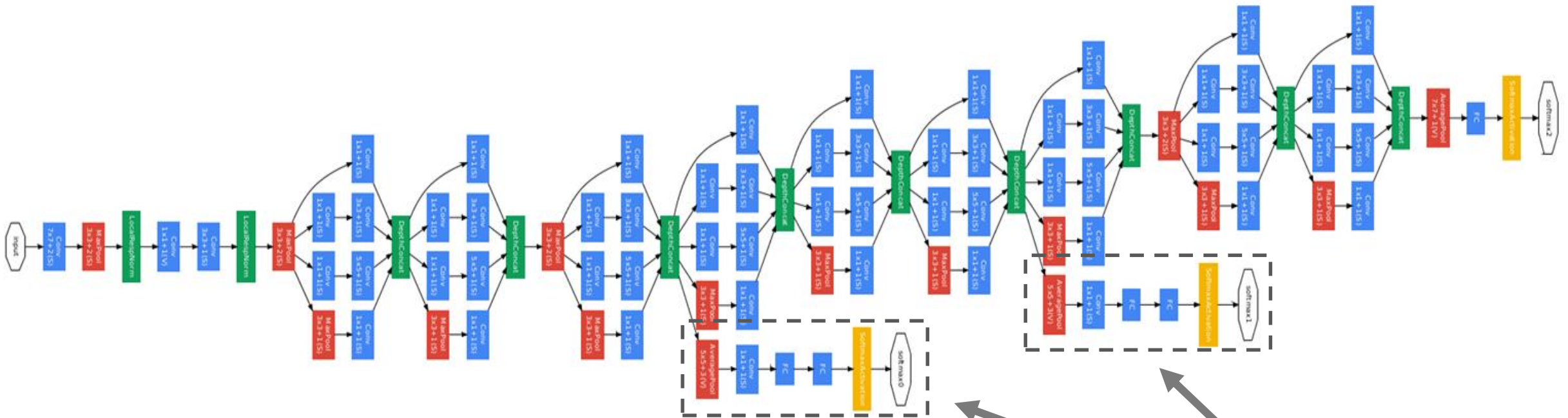


Agrupamento
pela Média
(average pooling)

Redução params

ConvNets – GoogLeNet

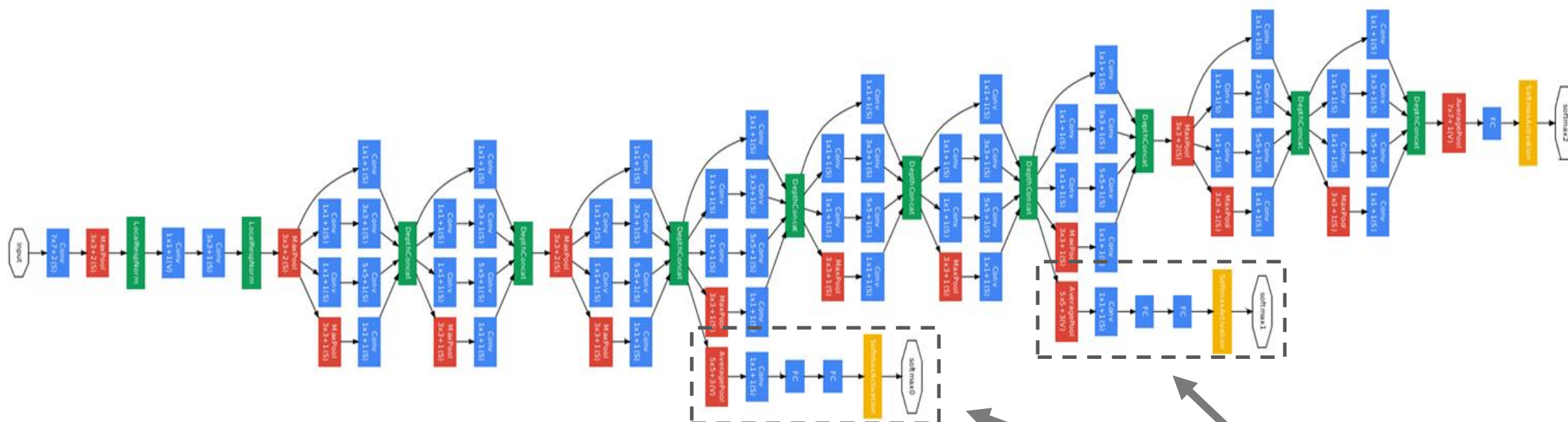
[Szegedy et al., 2014]



Classificadores
auxiliares

ConvNets – GoogLeNet

[Szegedy et al., 2014]



Classificadores
auxiliares

Facilita treinamento

ConvNets – GoogLeNet

[Szegedy et al., 2014]

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Vencedora ILSVRC 2014 – 6,7% de erro (top 5)

ConvNets – GoogLeNet

[Szegedy et al., 2014]

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Detalhes:

- Apenas ≈ 5M params!
(Remoção completa das camadas FC)

Vencedora ILSVRC 2014 – 6,7% de erro (top 5)

ConvNets – GoogLeNet

[Szegedy et al., 2014]

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Vencedora ILSVRC 2014 – 6,7% de erro (top 5)

Detalhes:

- Apenas $\approx 5\text{M}$ params!
(Remoção completa das camadas FC)

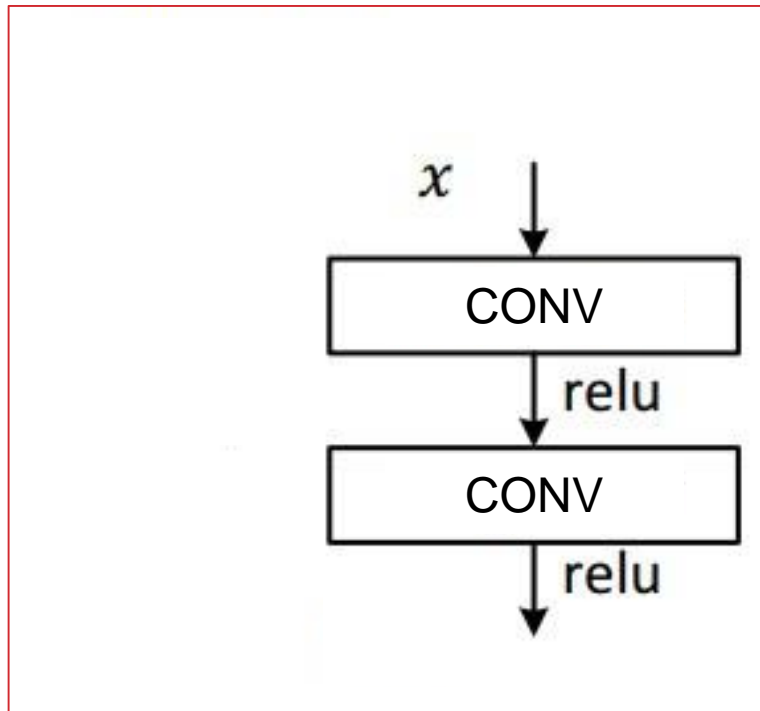
Comparada a AlexNet:

- 12× menos params
- 2× mais computação
- 6,67% de erro (vs. 16,4%)

ConvNets – ResNet

[He et al., 2015]

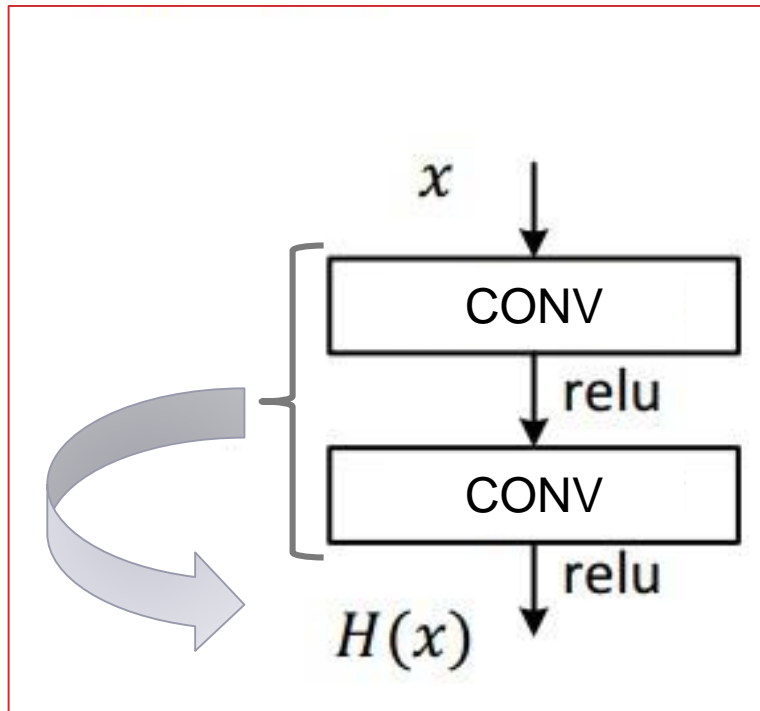
Rede Tradicional



ConvNets – ResNet

[He et al., 2015]

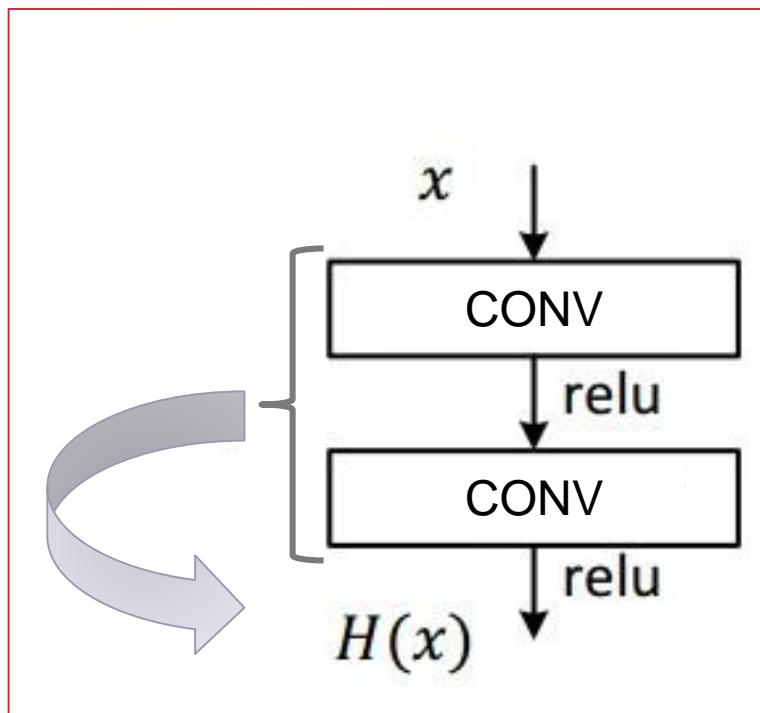
Rede Tradicional



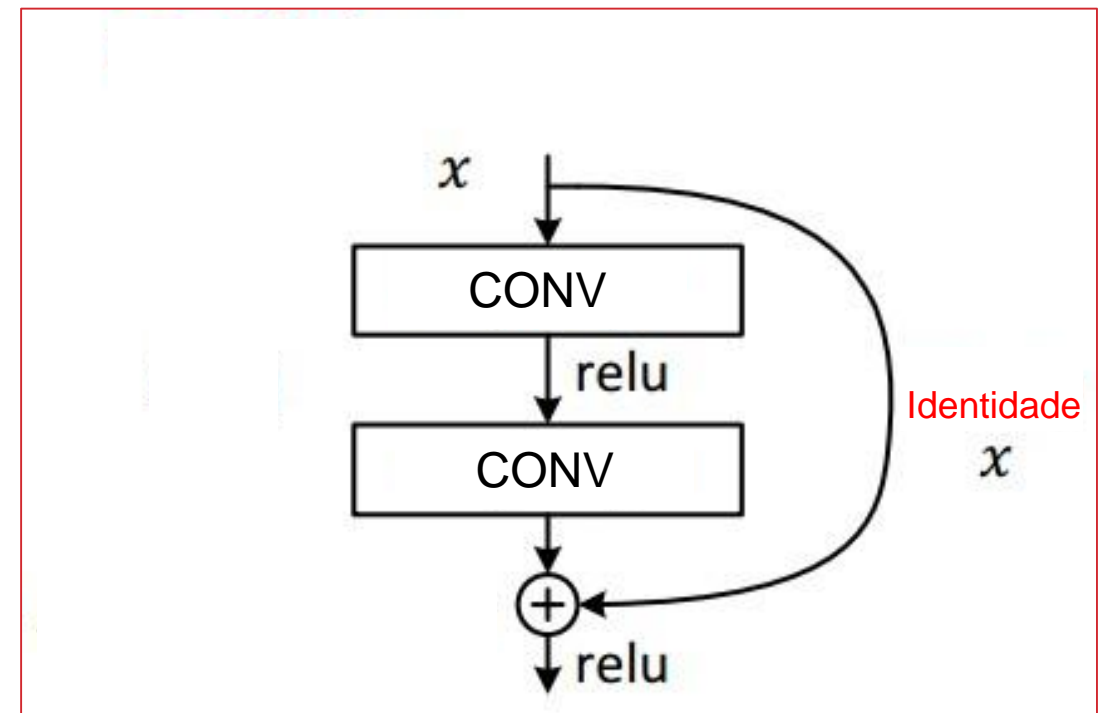
ConvNets – ResNet

[He et al., 2015]

Rede Tradicional



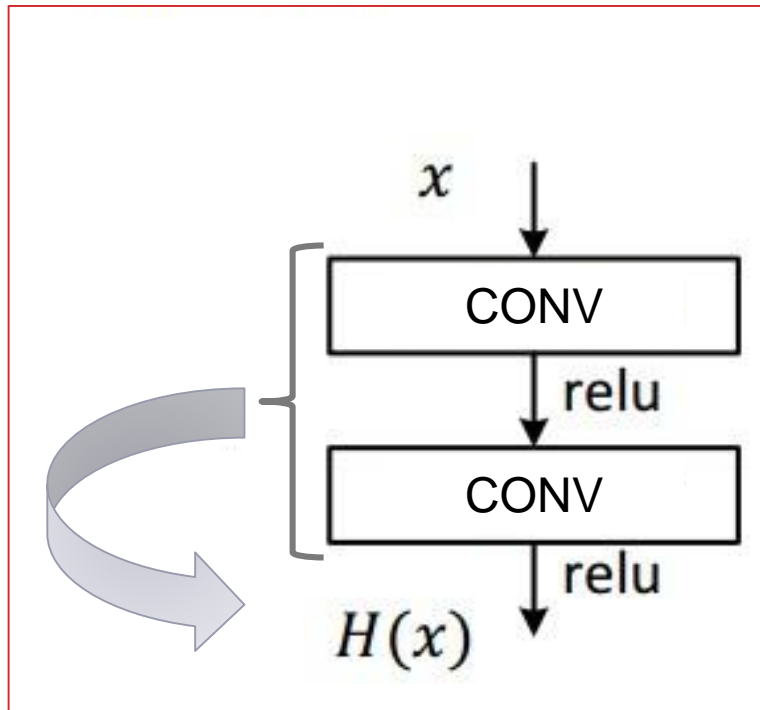
Rede Residual



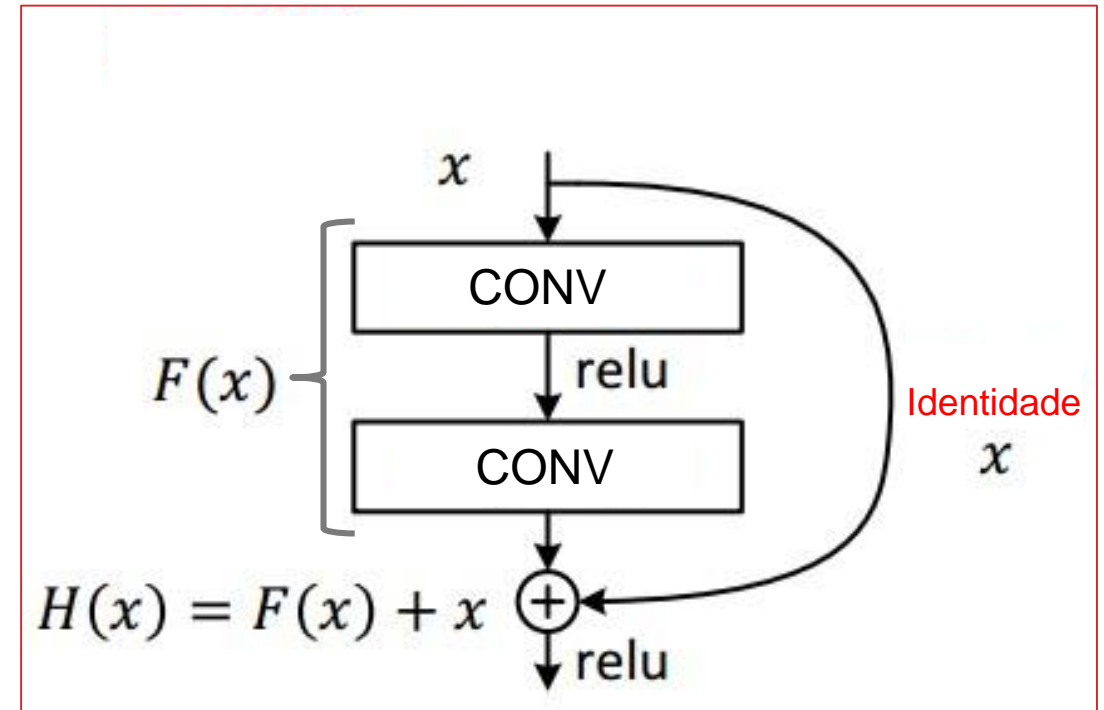
ConvNets – ResNet

[He et al., 2015]

Rede Tradicional



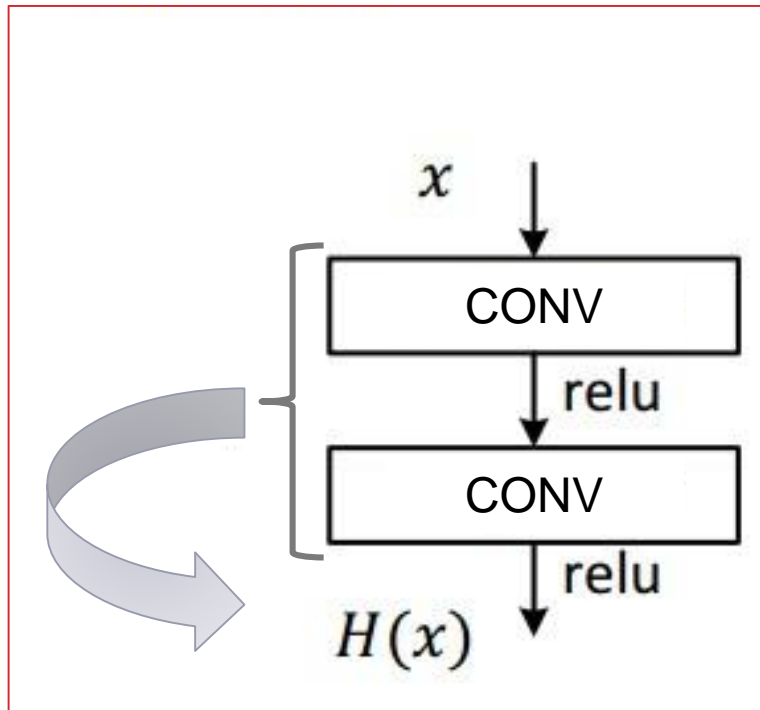
Rede Residual



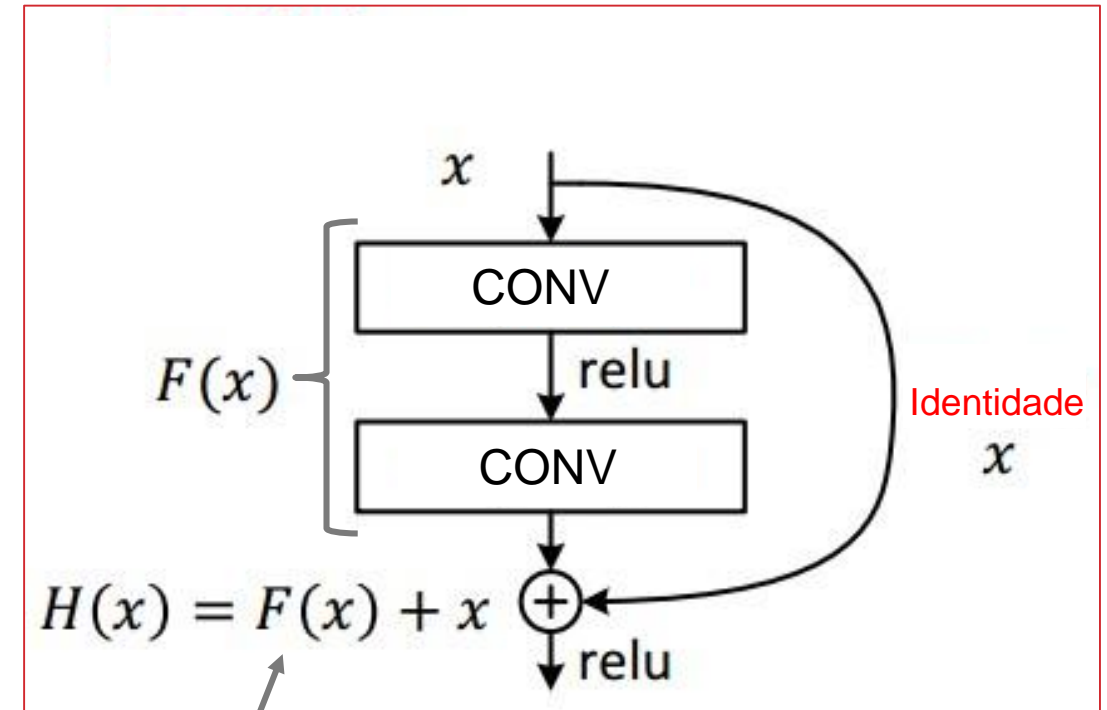
ConvNets – ResNet

[He et al., 2015]

Rede Tradicional



Rede Residual



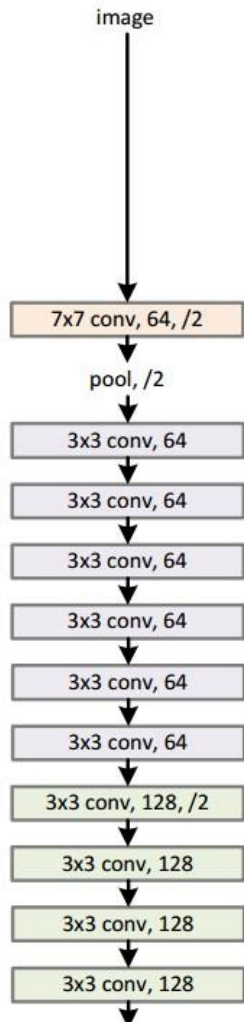
Resíduo

$$F(x) = H(x) - x$$

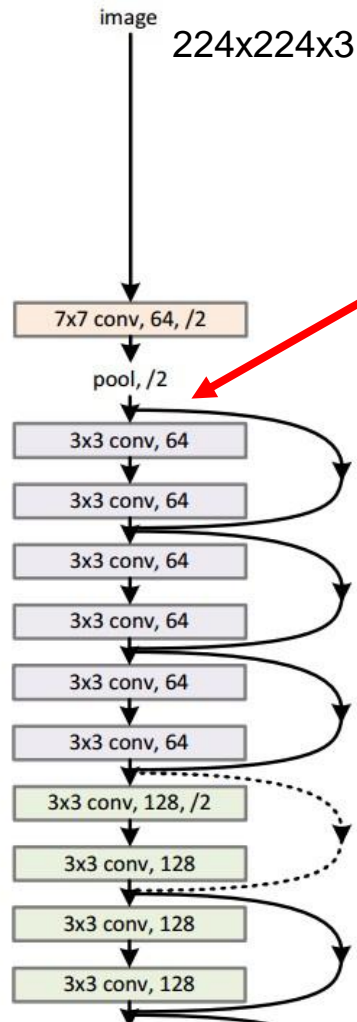
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual

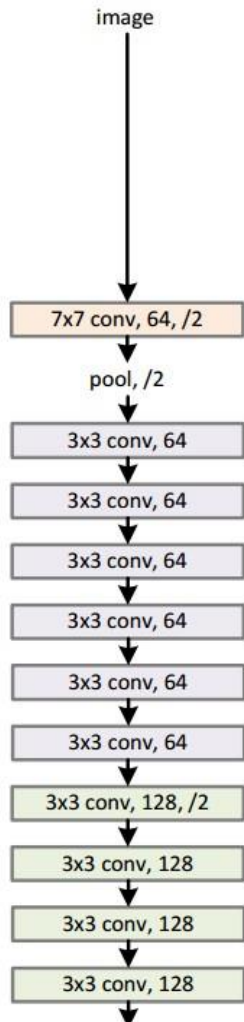


Dimensão espacial
apenas 56x56!

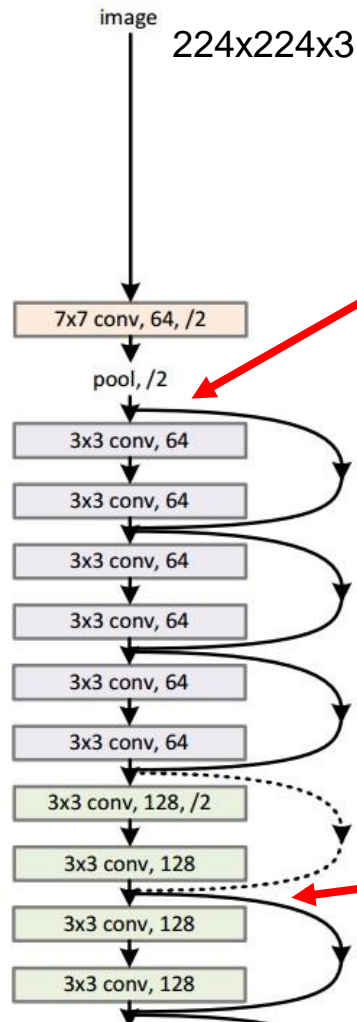
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



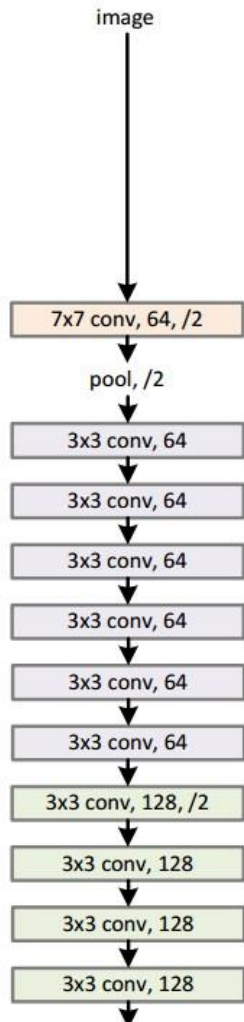
Dimensão espacial
apenas 56×56!

Dimensão espacial
apenas 28×28!

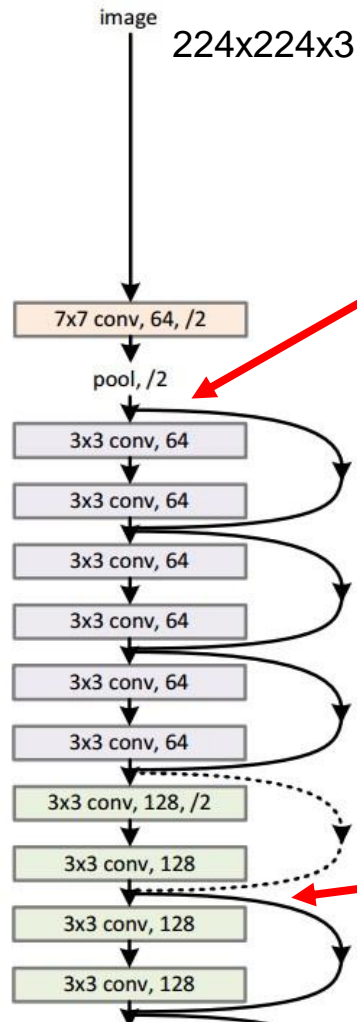
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



Dimensão espacial apenas 56×56!

Dimensão espacial apenas 28×28!

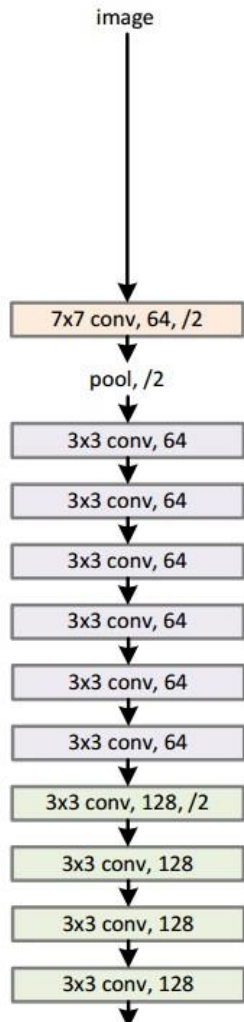
Detalhes:

- Normalização em lote após cada CONV

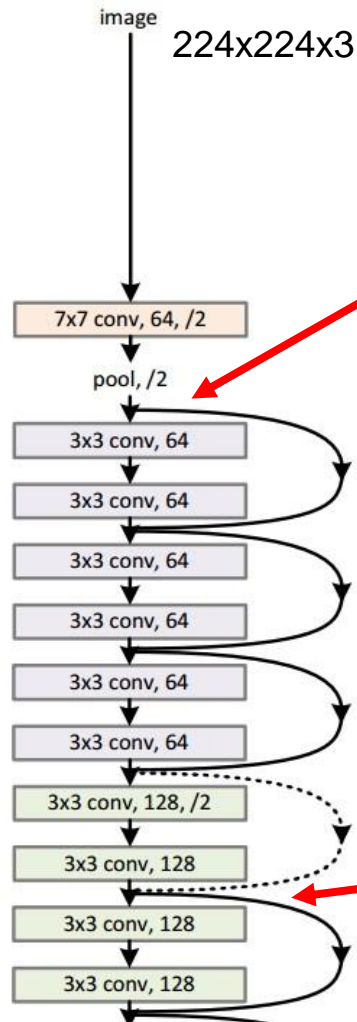
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



Dimensão espacial apenas 56×56!

Dimensão espacial apenas 28×28!

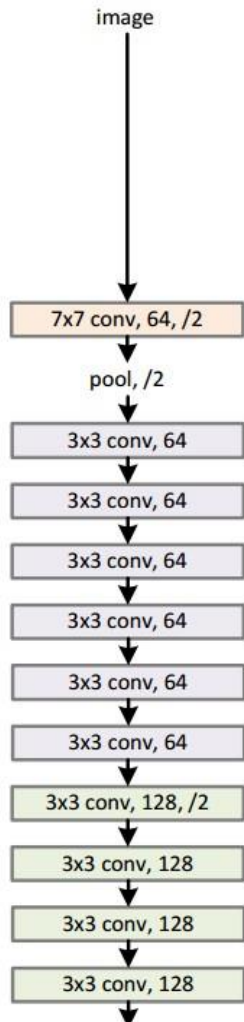
Detalhes:

- Normalização em lote após cada CONV
- Inicialização modificada (Xavier + /2)

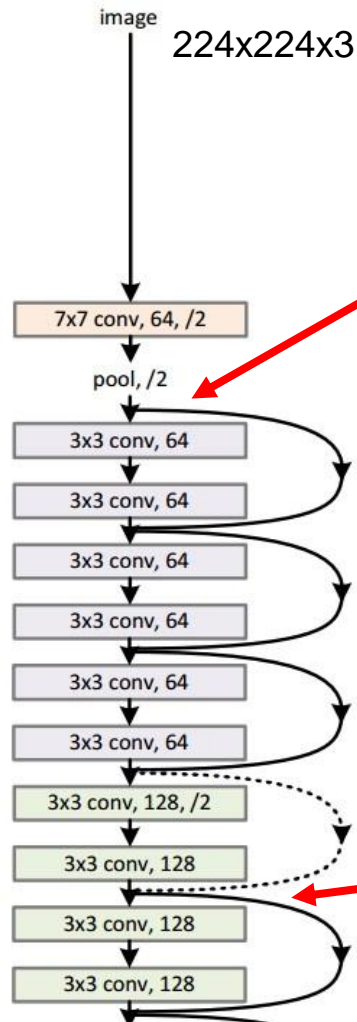
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



Dimensão espacial apenas 56×56!

Dimensão espacial apenas 28×28!

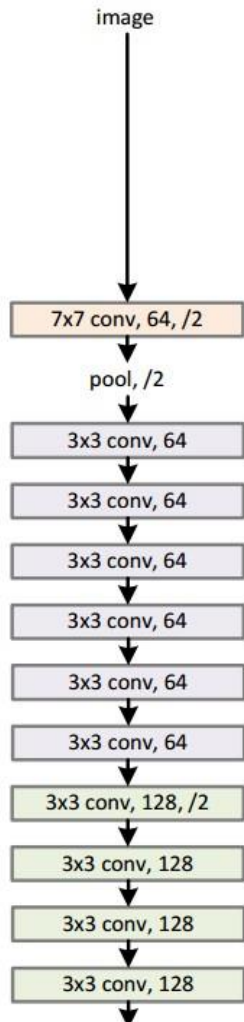
Detalhes:

- Normalização em lote após cada CONV
- Inicialização modificada (Xavier + /2)
- SGD+*Momentum* (0,9)

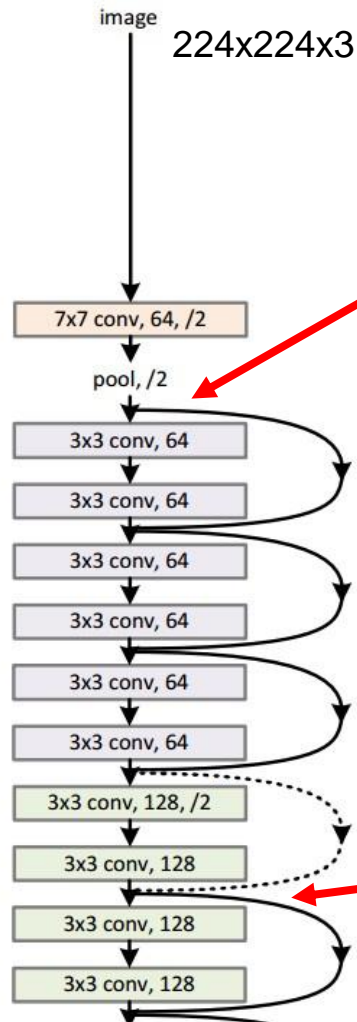
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



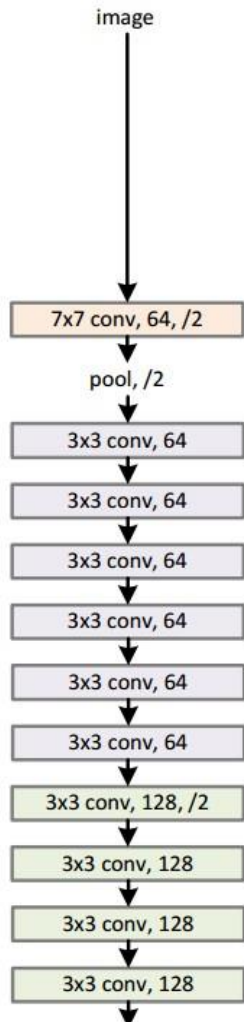
Detalhes:

- Normalização em lote após cada CONV
- Inicialização modificada (Xavier + /2)
- SGD+*Momentum* (0,9)
- Tx aprendizado = 10^{-1} , dividida por 10 qdo erro de validação para de reduzir

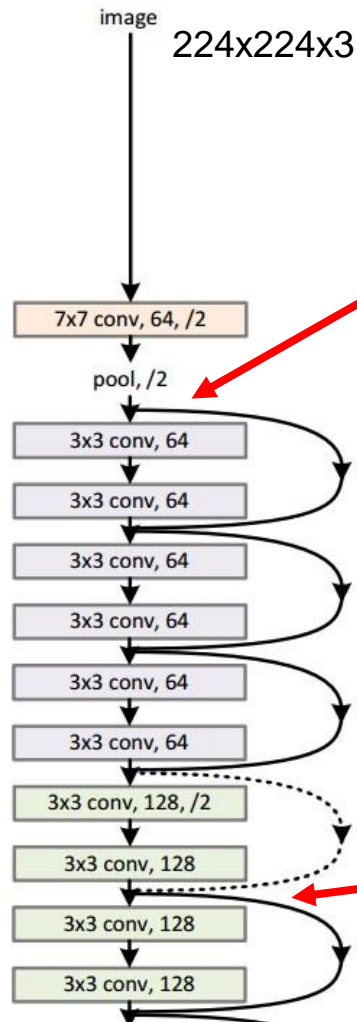
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



Dimensão espacial
apenas 56×56!

Dimensão espacial
apenas 28×28!

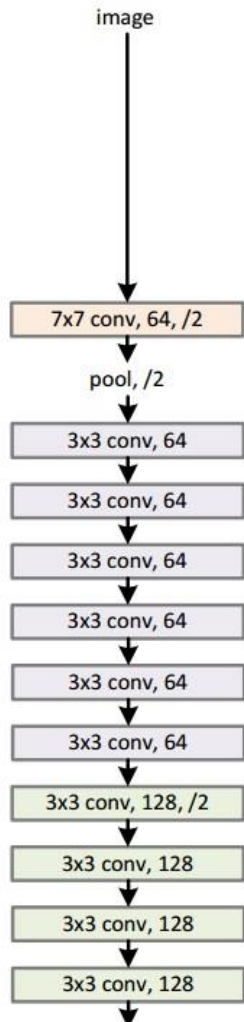
Detalhes:

- Normalização em lote após cada CONV
- Inicialização modificada (Xavier + /2)
- SGD+*Momentum* (0,9)
- Tx aprendizado = 10^{-1} , dividida por 10 qdo erro de validação para de reduzir
- Tamanho do *minibatch* = 256

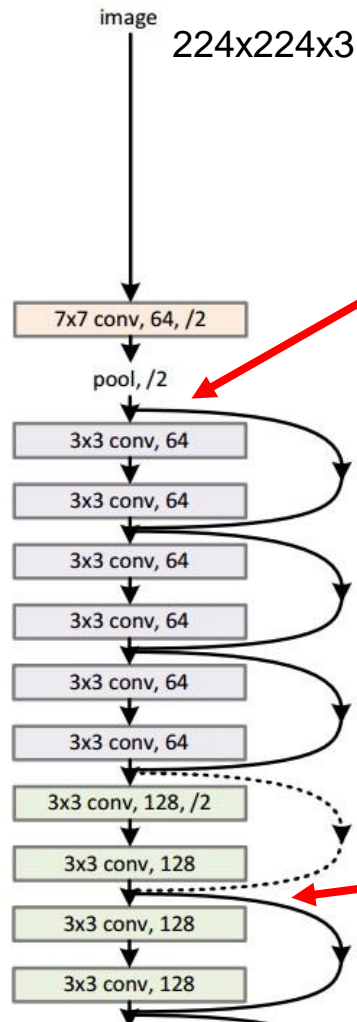
ConvNets – ResNet

[He et al., 2015]

34-layer plain



34-layer residual



Dimensão espacial apenas 56×56!

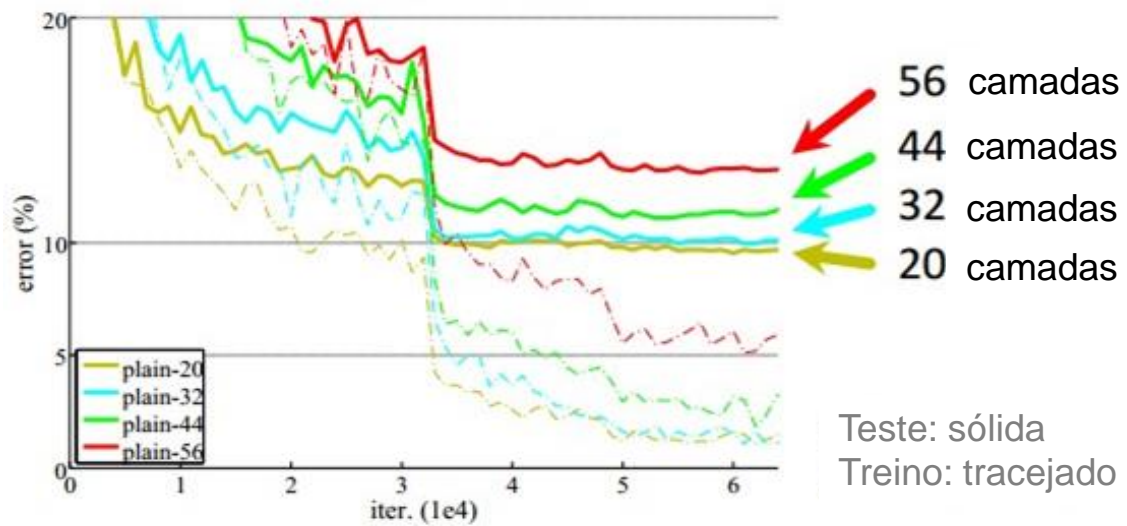
Dimensão espacial apenas 28×28!

Detalhes:

- Normalização em lote após cada CONV
- Inicialização modificada (Xavier + /2)
- SGD+*Momentum* (0,9)
- Tx aprendizado = 10^{-1} , dividida por 10 qdo erro de validação para de reduzir
- Tamanho do *minibatch* = 256
- *Dropout* não é usado

ConvNets – ResNet

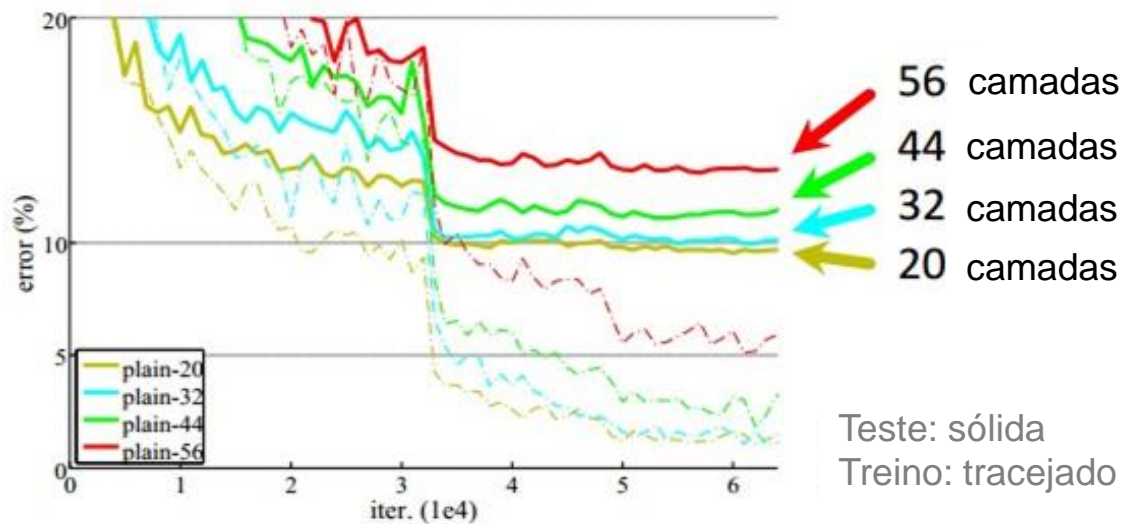
Experimentos com o conjunto de dados CIFAR-10



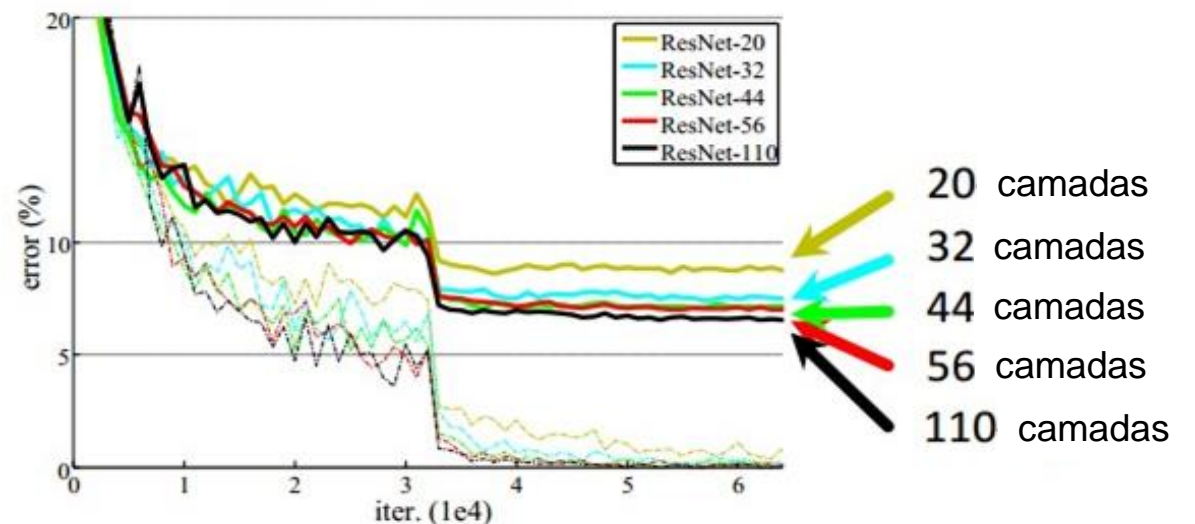
Rede Tradicional

ConvNets – ResNet

Experimentos com o conjunto de dados CIFAR-10



Rede Tradicional

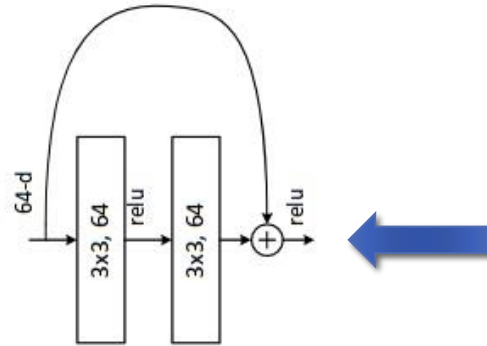


Rede Residual

ConvNets – ResNet

[He et al., 2015]

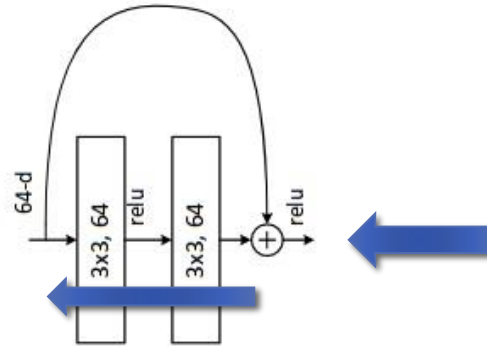
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

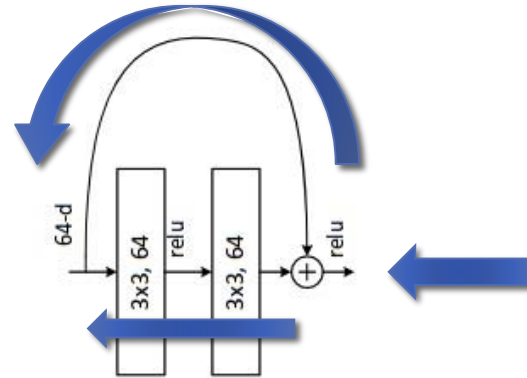
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

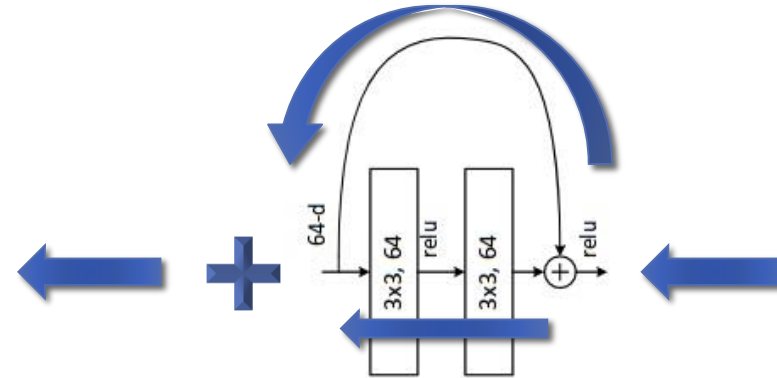
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

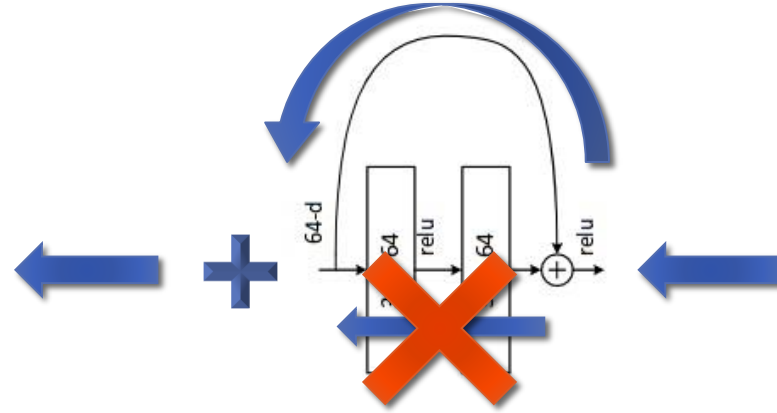
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

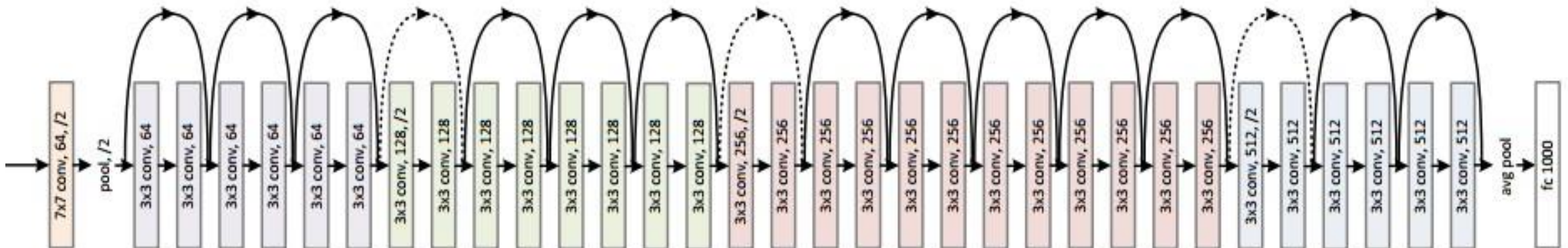
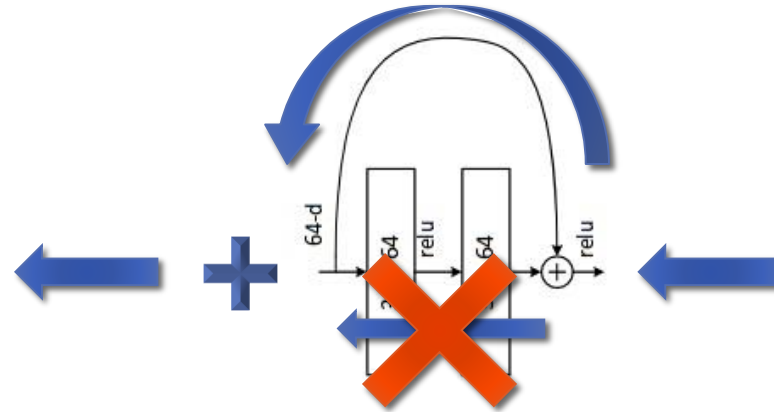
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

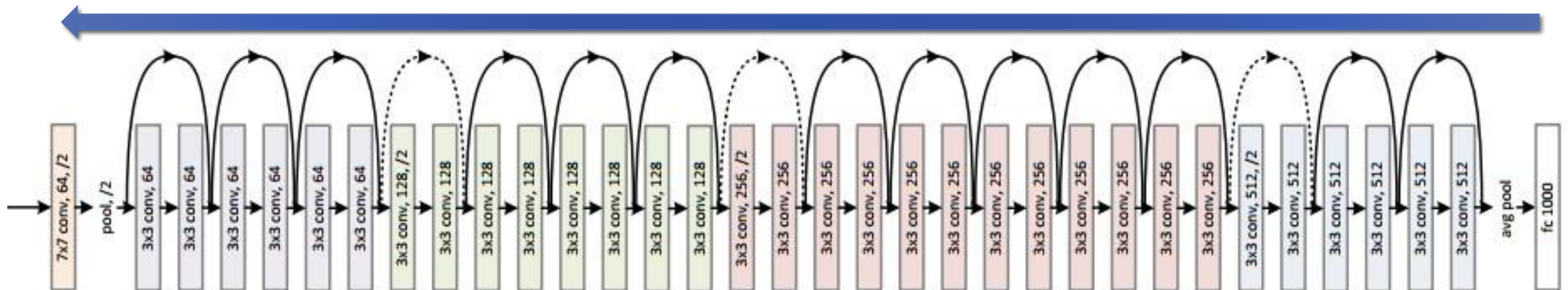
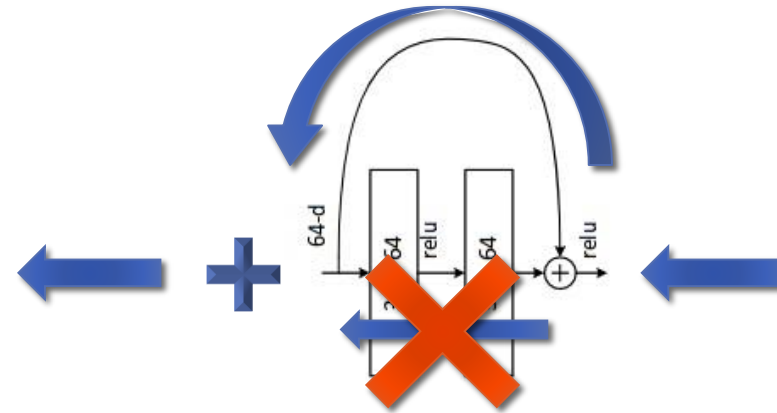
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

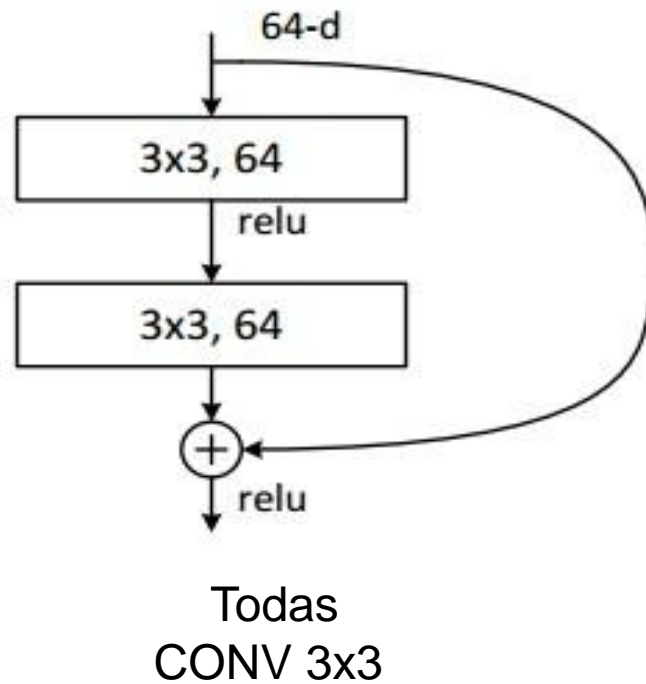
Fluxo de Gradientes



ConvNets – ResNet

[He et al., 2015]

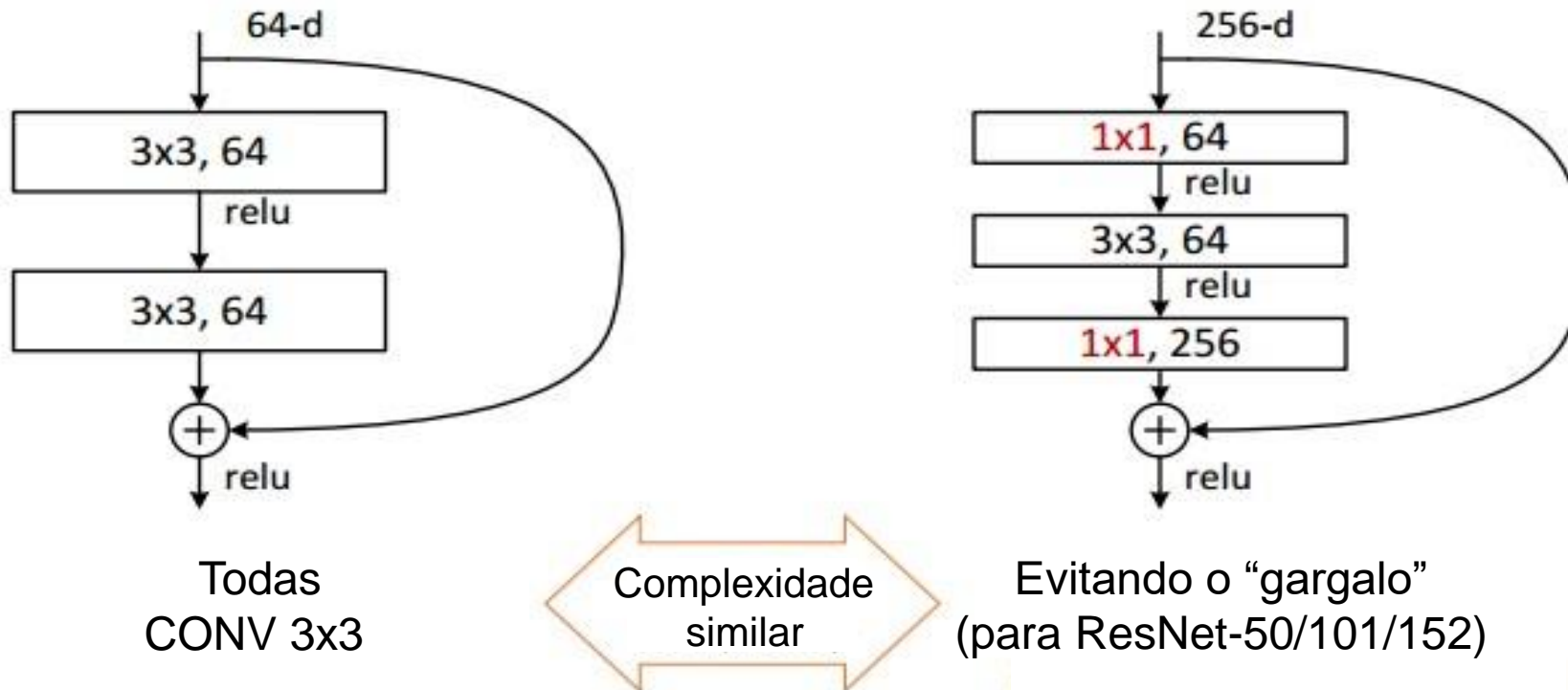
Lidando com redes de 50+ camadas



ConvNets – ResNet

[He et al., 2015]

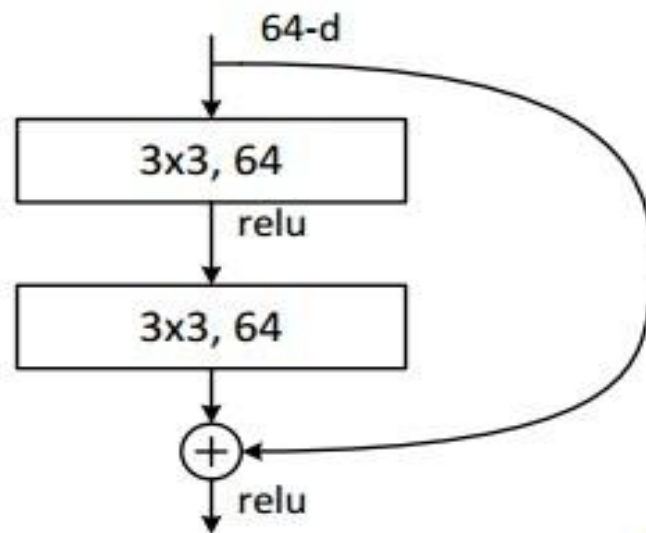
Lidando com redes de 50+ camadas



ConvNets – ResNet

[He et al., 2015]

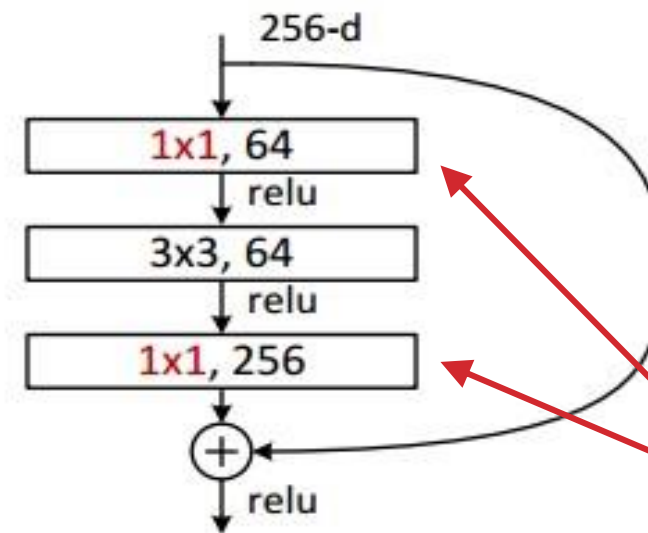
Lidando com redes de 50+ camadas



Todas
CONV 3x3



Complexidade
similar

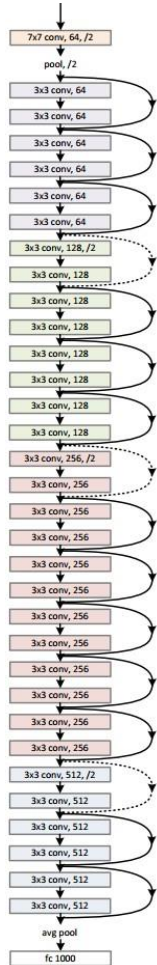


Evitando o “gargalo”
(para ResNet-50/101/152)

Truque similar
a GoogLeNet

ConvNets – ResNet

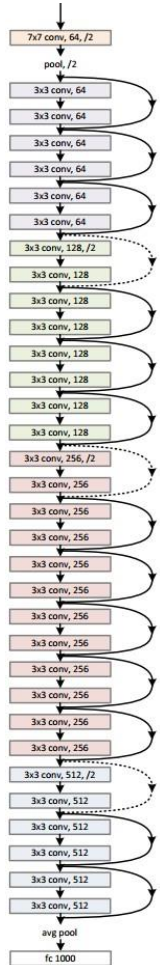
[He et al., 2015]



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

ConvNets – ResNet

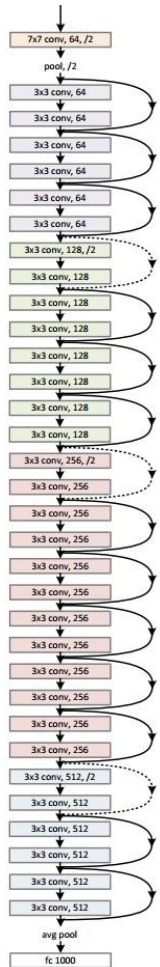
[He et al., 2015]



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

ConvNets – ResNet

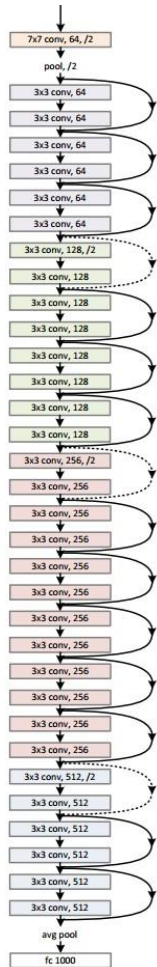
[He et al., 2015]



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

ConvNets – ResNet

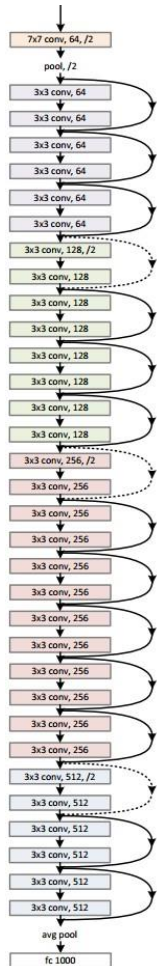
[He et al., 2015]



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

ConvNets – ResNet

[He et al., 2015]

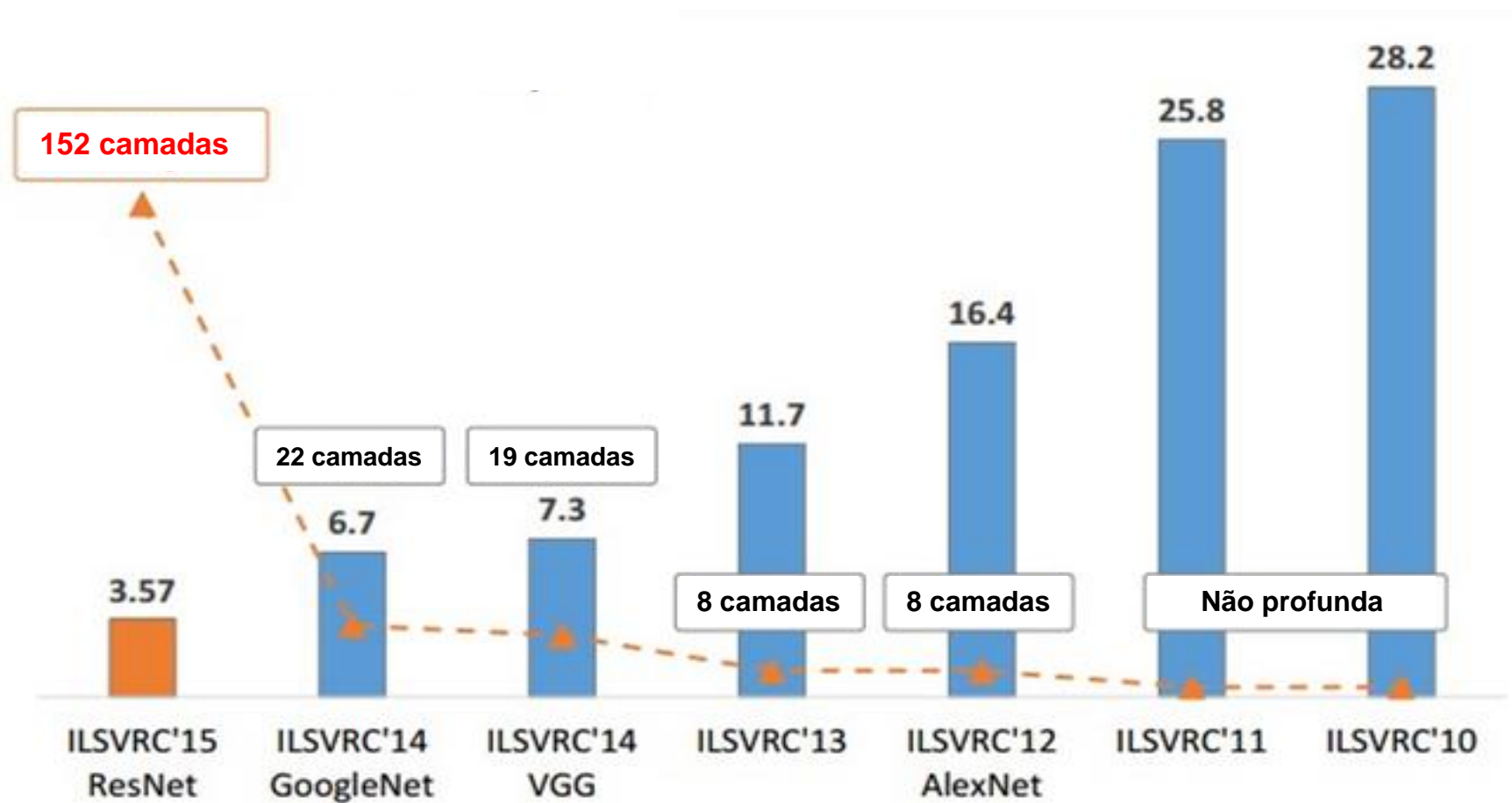


layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

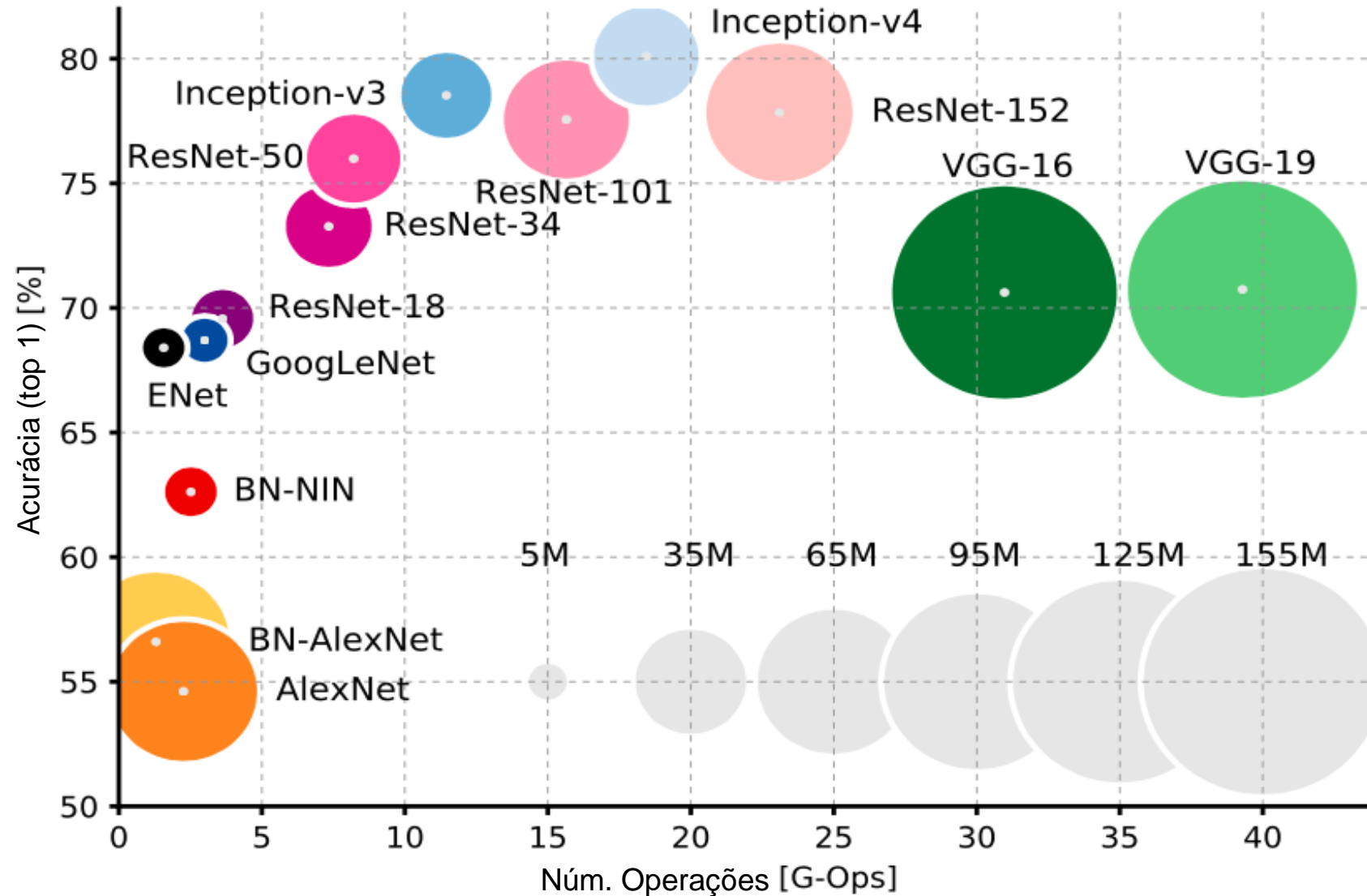
VGG16 – $15,3 \times 10^9$
VGG19 – $19,6 \times 10^9$

Vencedora ILSVRC 2015 – 3,6% de erro (top 5)

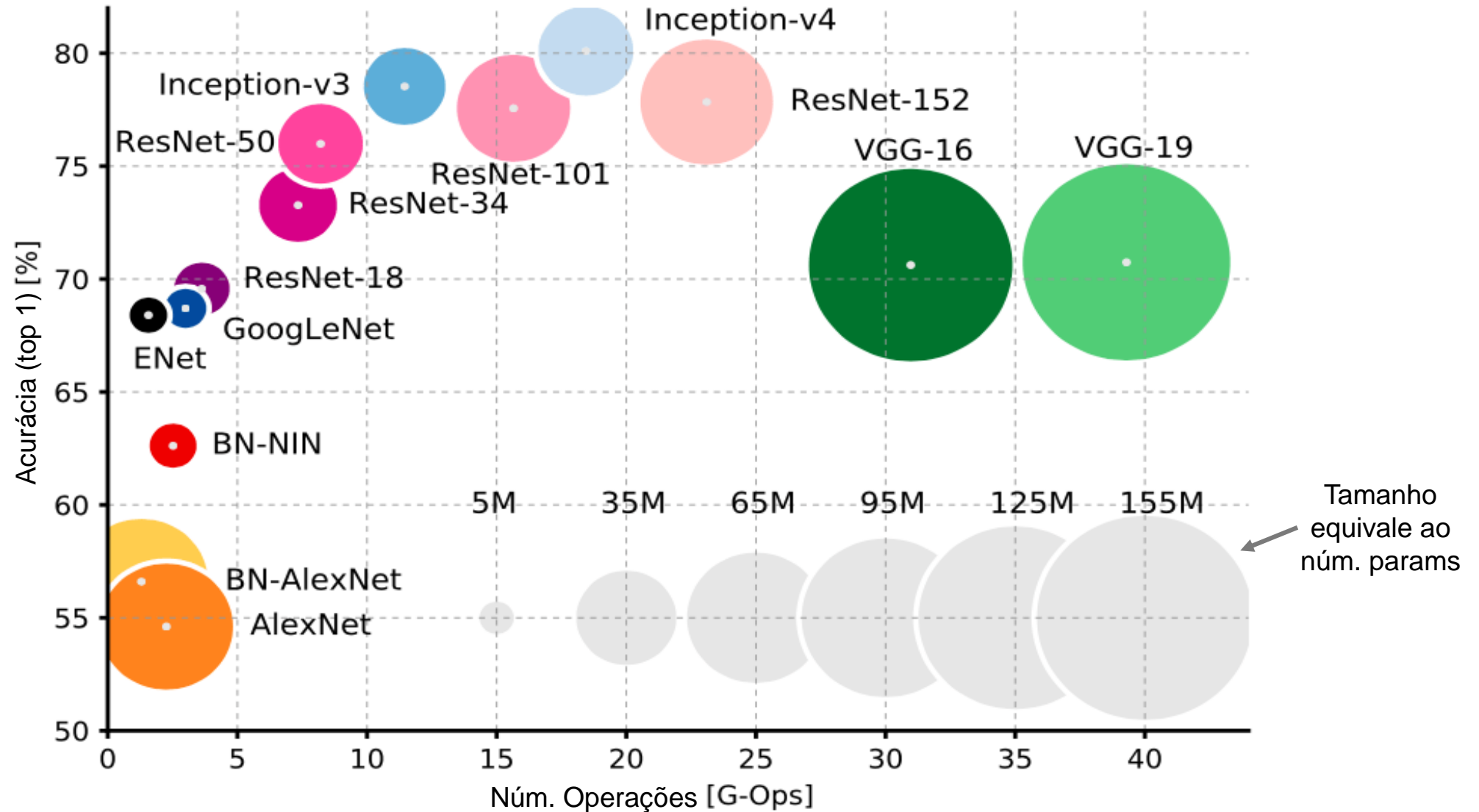
ConvNets – Análise Comparativa



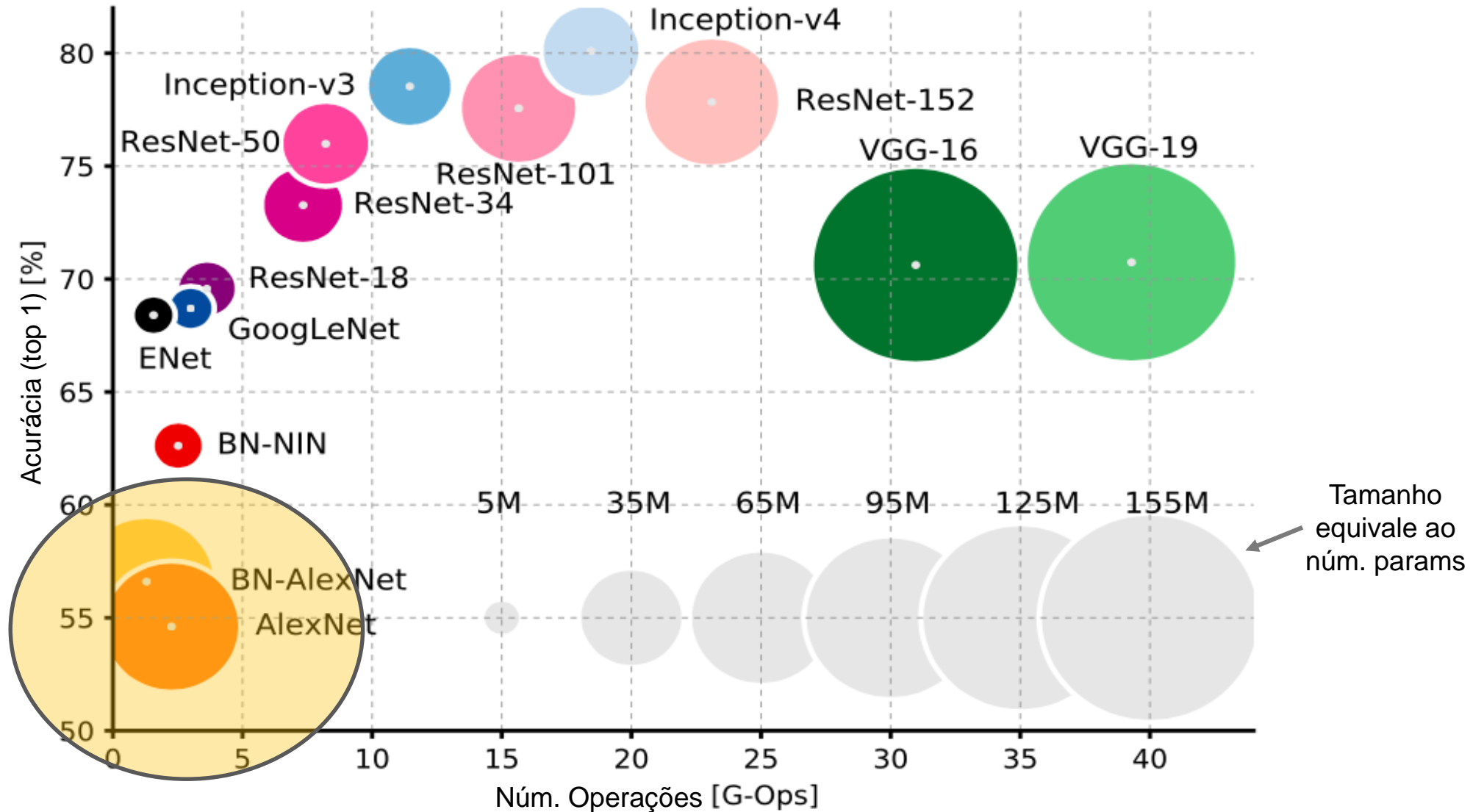
ConvNets – Análise Comparativa



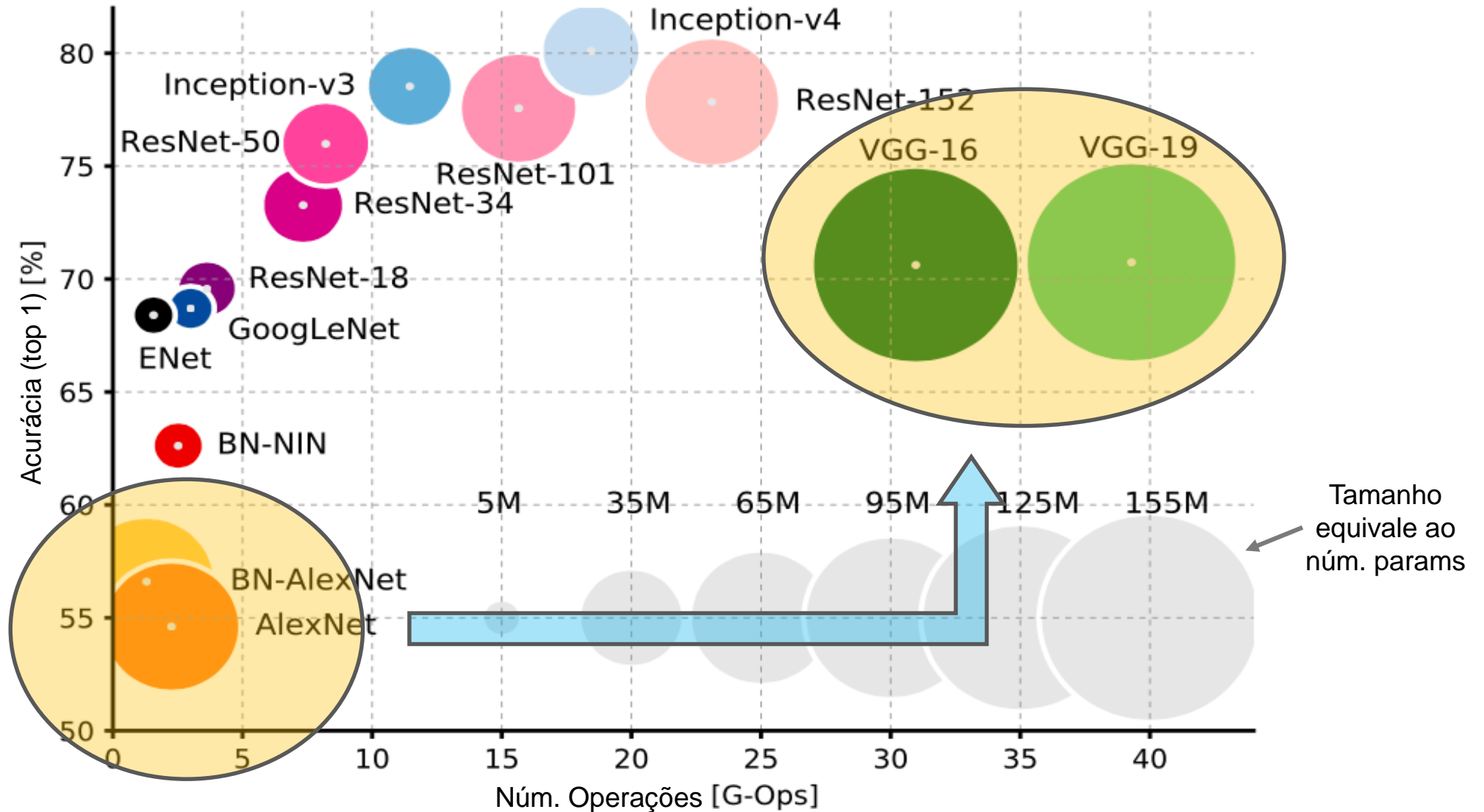
ConvNets – Análise Comparativa



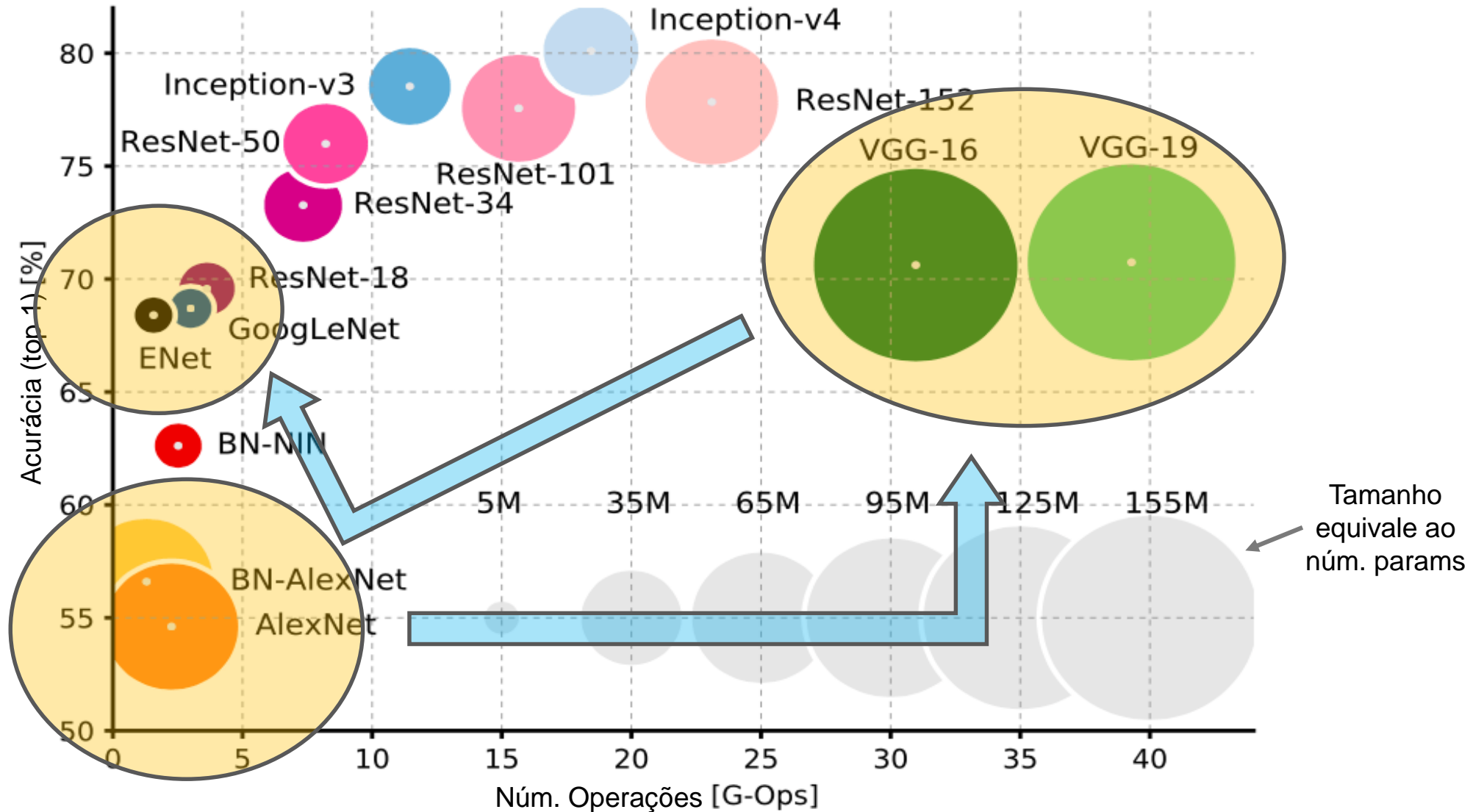
ConvNets – Análise Comparativa



ConvNets – Análise Comparativa



ConvNets – Análise Comparativa



ConvNets – Análise Comparativa

