



MINERAÇÃO DE DADOS COMPLEXOS

Curso de aperfeiçoamento



INF-0618

Tópicos em Aprendizado de Máquina II

Aula 7 - Aplicações avançadas

Detecção de objetos / Transferência de estilo

Profa. Fernanda Andaló

2018

Instituto de Computação - Unicamp

Slides baseados em:
<http://cs231n.stanford.edu/>
<https://www.coursera.org/learn/convolutional-neural-networks>

Roteiro

Detecção e segmentação de objetos

Transferência de estilo

Detecção e segmentação de objetos

Detecção e segmentação de objetos

Até agora: classificação de objetos



This image is CC0 public domain

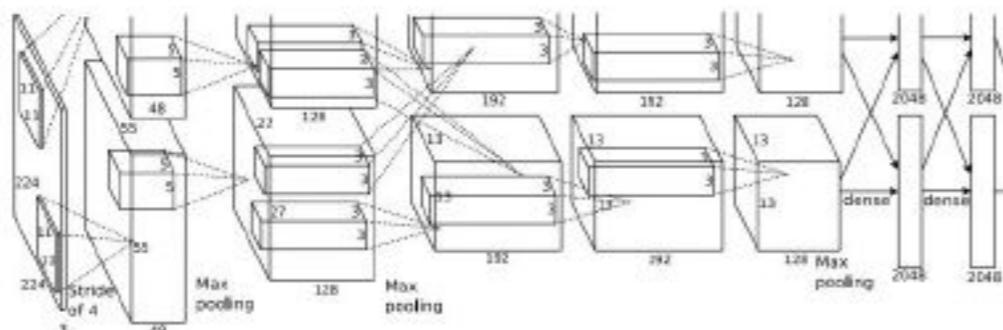


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Vector:
4096

Fully-Connected:
4096 to 1000

Class Scores
Cat: 0.9
Dog: 0.05
Car: 0.01
...

Detecção e segmentação de objetos

Outras tarefas

Segmentação
semântica
(apenas pixels)



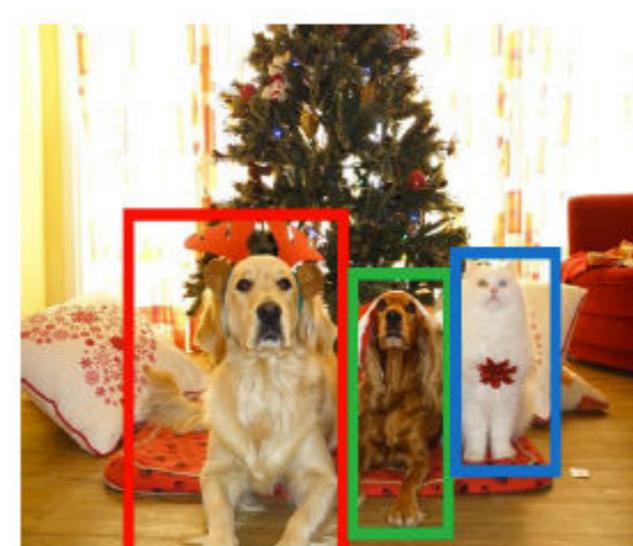
GRASS, CAT,
TREE, SKY

Classificação +
Localização
(objeto único)



CAT

Detecção
(objetos múltiplos)



DOG, DOG, CAT

Segmentação de
instâncias
(objetos múltiplos)



DOG, DOG, CAT

Detecção e segmentação de objetos

Segmentação
semântica
(apenas pixels)



**GRASS, CAT,
TREE, SKY**

Classificação +
Localização
(objeto único)



CAT

Detecção
(objetos múltiplos)



DOG, DOG, CAT

Segmentação de
instâncias
(objetos múltiplos)

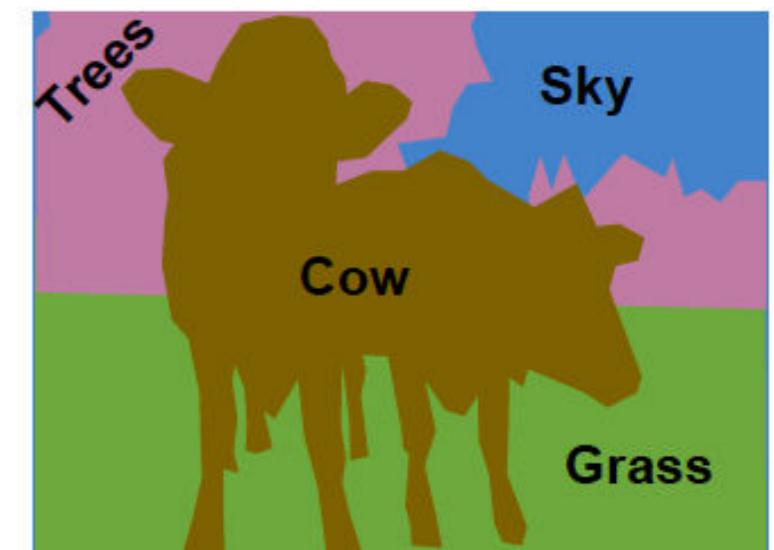
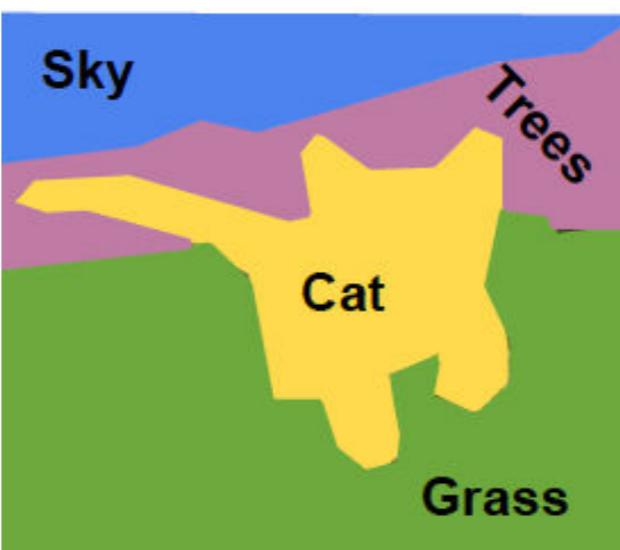


DOG, DOG, CAT

Detecção e segmentação de objetos

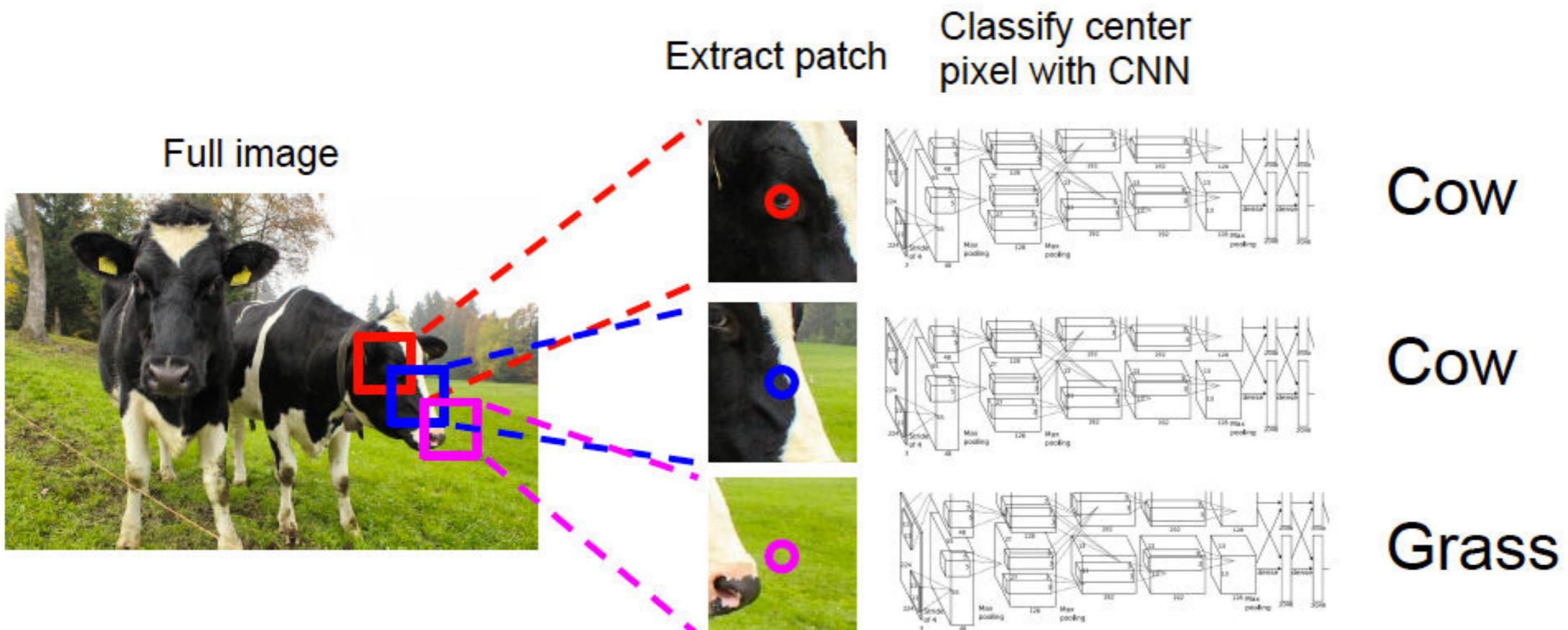
Segmentação semântica

- Rotular cada pixel da imagem com um rótulo de uma classe.
- Não diferenciar instâncias, apenas classificar cada pixel.



Detecção e segmentação de objetos

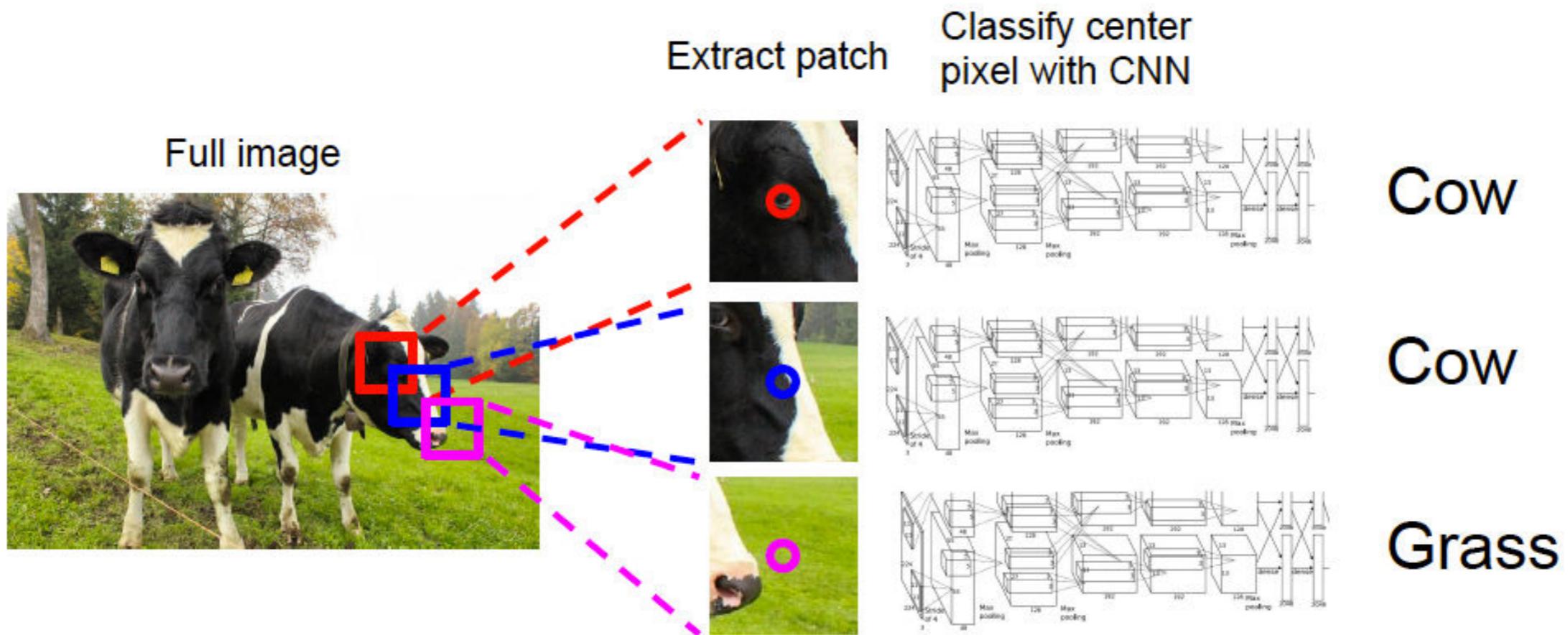
Segmentação semântica: janela deslizante



Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

Detecção e segmentação de objetos

Segmentação semântica: janela deslizante



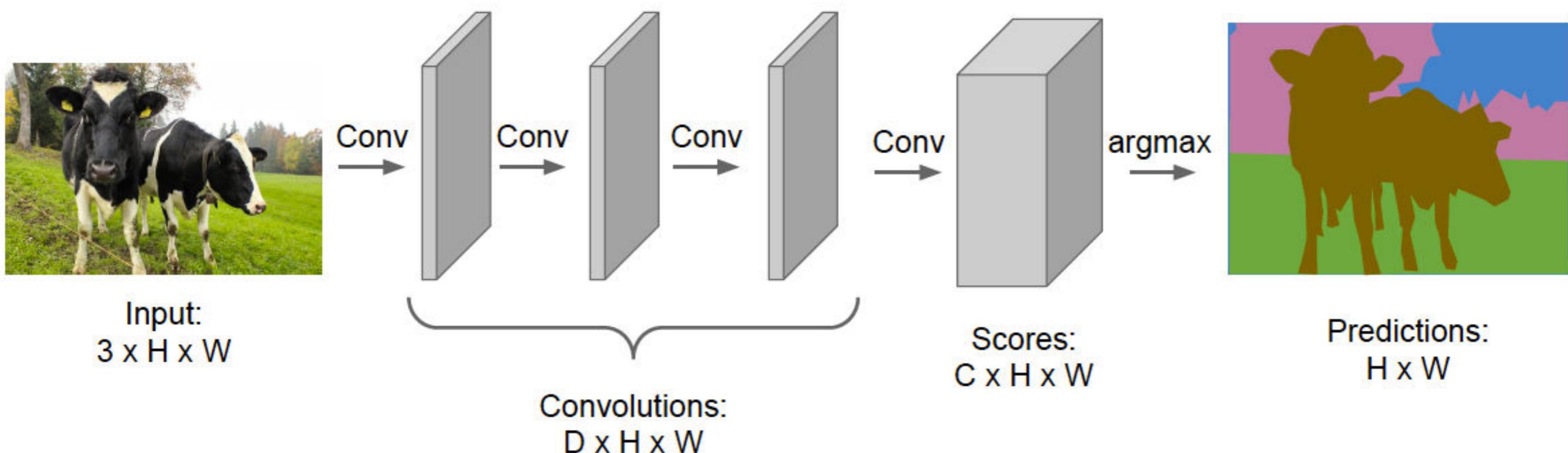
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

Problema! Muito ineficiente! Não está usando features compartilhadas em patches que se sobrepõem.

Detecção e segmentação de objetos

Segmentação semântica: fully convolutional

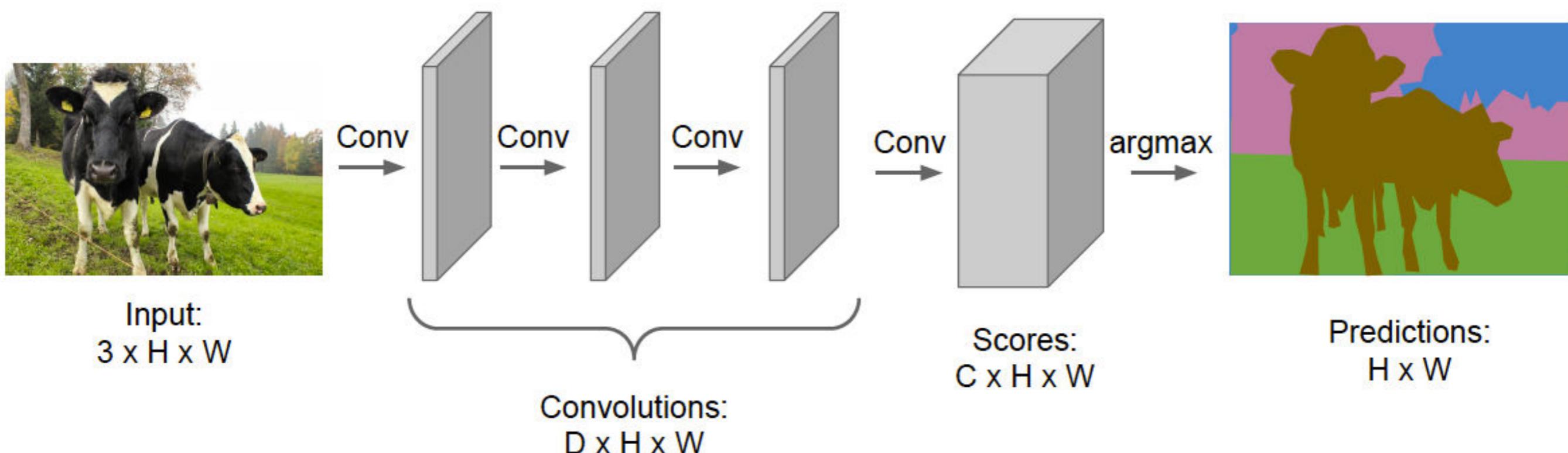
Projete uma rede com muitas camadas convolucionais para fazer previsões para todos os pixels ao mesmo tempo.



Detecção e segmentação de objetos

Segmentação semântica: fully convolutional

Projete uma rede com muitas camadas convolucionais para fazer previsões para todos os pixels ao mesmo tempo.



Problema! Convoluções na resolução original da imagem
são bem caras.

Detecção e segmentação de objetos

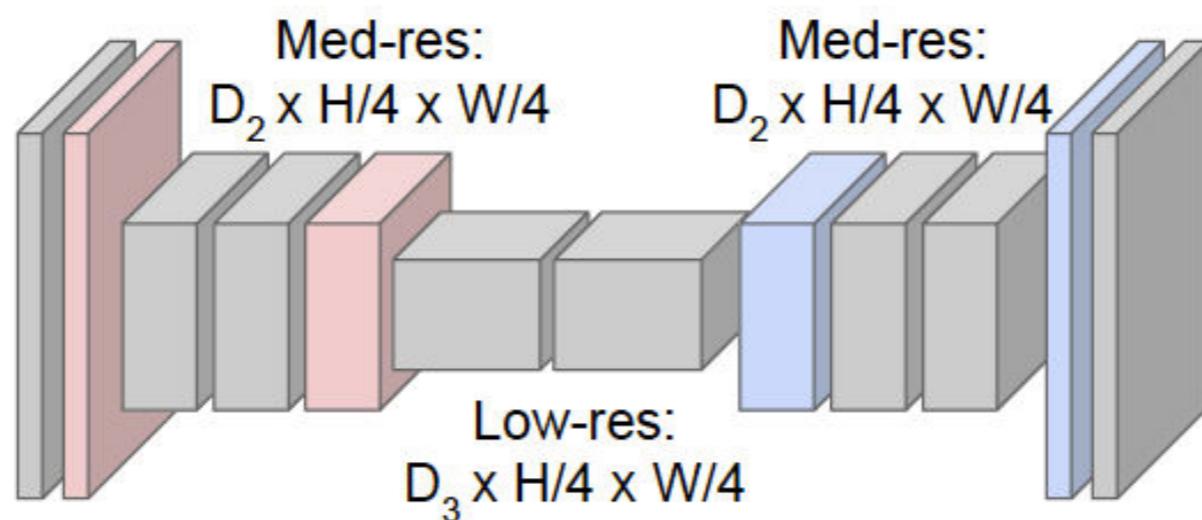
Segmentação semântica: fully convolutional

Projete uma rede com muitas camadas convolucionais, com **downsampling** e **upsampling** dentro da rede.



Input:
 $3 \times H \times W$

High-res:
 $D_1 \times H/2 \times W/2$



High-res:
 $D_1 \times H/2 \times W/2$



Predictions:
 $H \times W$

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Detecção e segmentação de objetos

Segmentação semântica: fully convolutional

Downsampling:

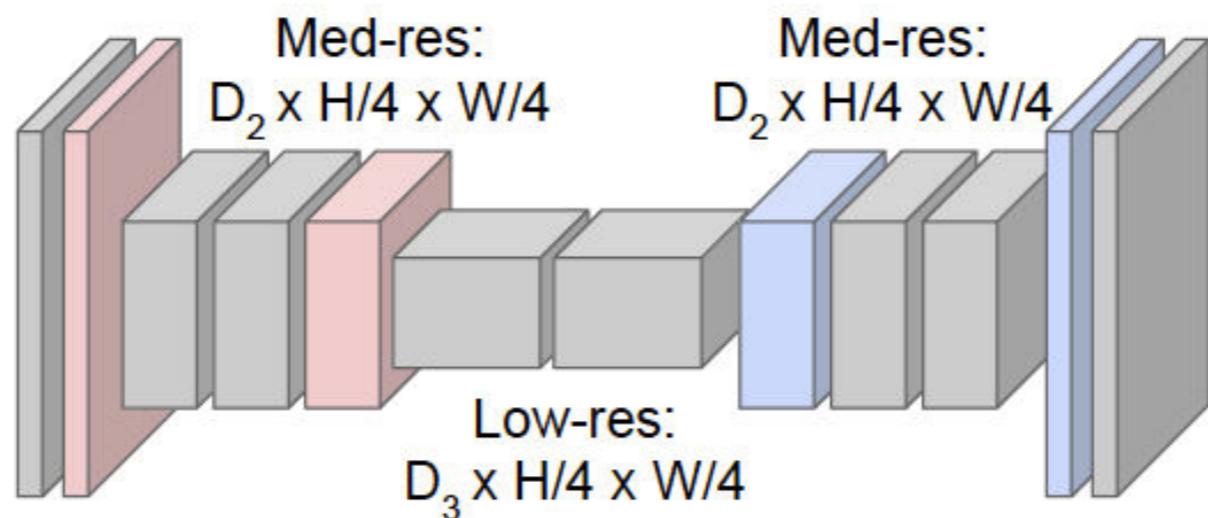
pooling,
convolução com
stride

Projete uma rede com muitas camadas convolucionais,
com **downsampling** e **upsampling** dentro da rede.



Input:
 $3 \times H \times W$

High-res:
 $D_1 \times H/2 \times W/2$



Predictions:
 $H \times W$

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015

Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Detecção e segmentação de objetos

Segmentação semântica: fully convolutional

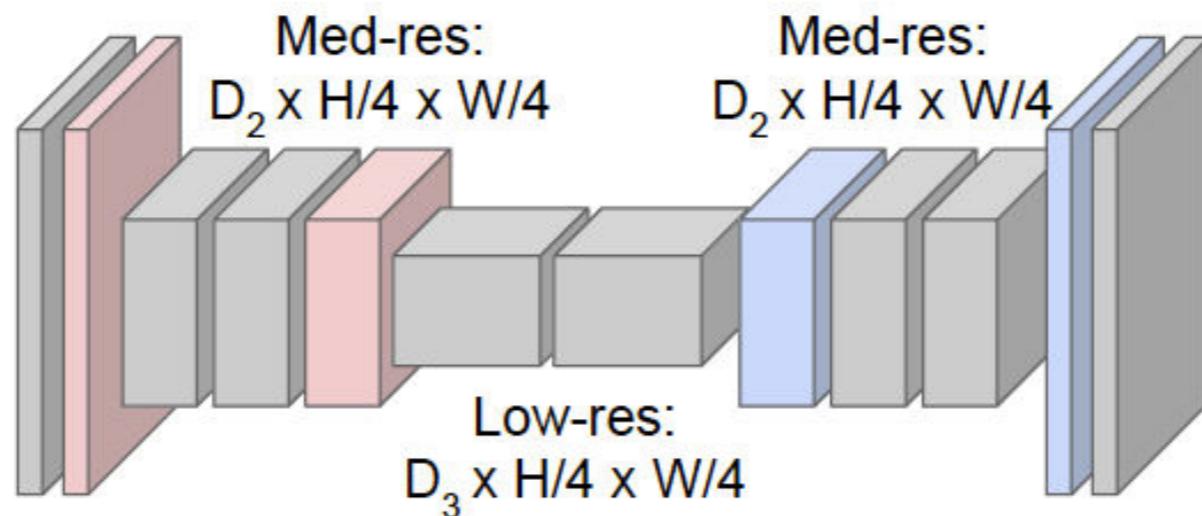
Downsampling:
pooling,
convolução com
stride



Projete uma rede com muitas camadas convolucionais,
com **downsampling** e **upsampling** dentro da rede.

Input:
 $3 \times H \times W$

High-res:
 $D_1 \times H/2 \times W/2$



Upsampling: ???



Predictions:
 $H \times W$

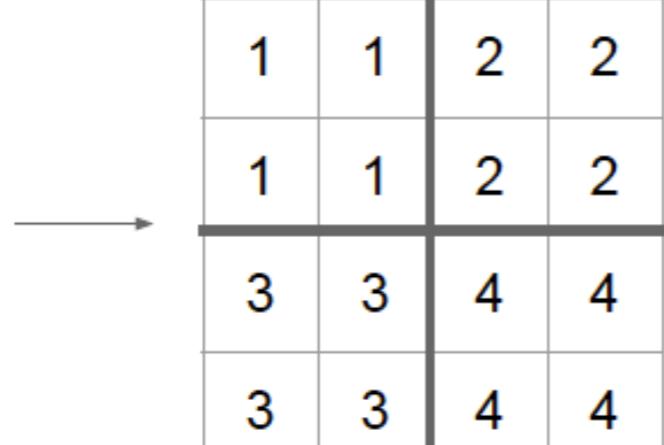
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Detecção e segmentação de objetos

Upsampling na rede: sem aprendizado

Nearest neighbor

1	2
3	4



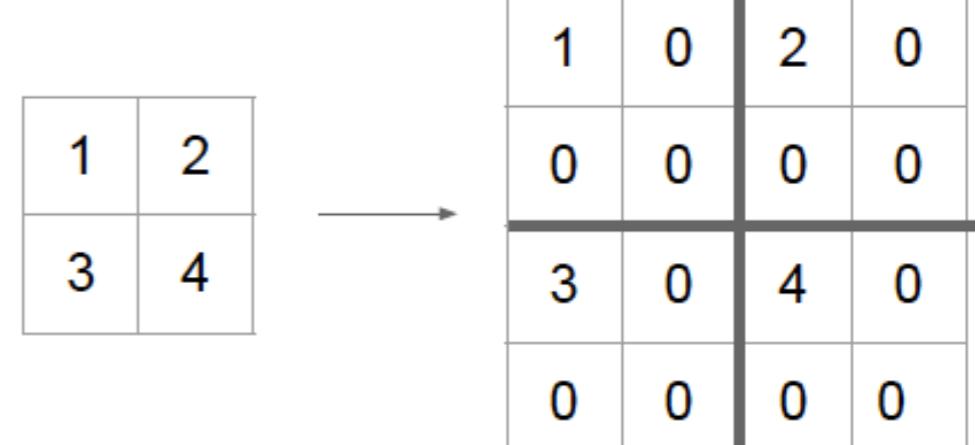
1	1	2	2
1	1	2	2
3	3	4	4
3	3	4	4

Input: 2 x 2

Output: 4 x 4

“Bed of nails”

1	2
3	4



1	0	2	0
0	0	0	0
3	0	4	0
0	0	0	0

Input: 2 x 2

Output: 4 x 4

Detecção e segmentação de objetos

Upsampling na rede: sem aprendizado

Max unpooling

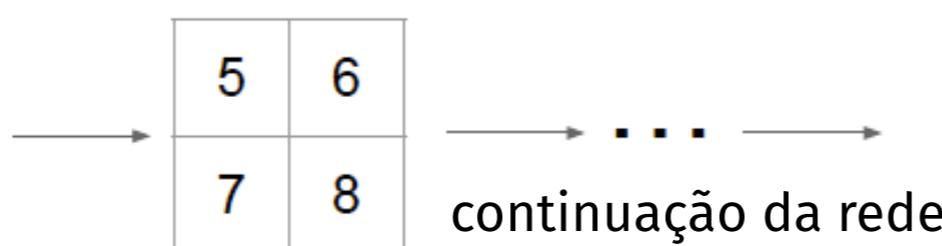
Max Pooling

Lembrar qual elemento foi max

1	2	6	3
3	5	2	1
1	2	2	1
7	3	4	8

Input: 4 x 4

Output: 2 x 2



Max unpooling

Usar posições da camada de pooling

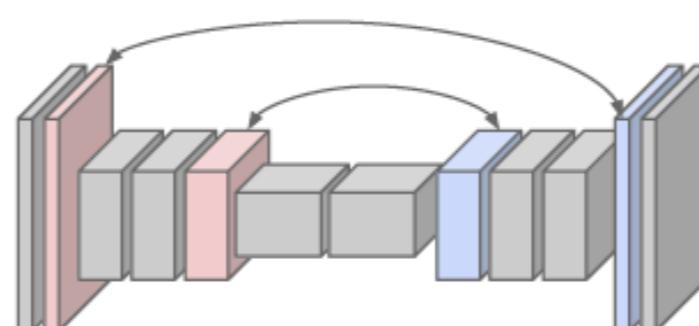
1	2
3	4

Input: 2 x 2

0	0	2	0
0	1	0	0
0	0	0	0
3	0	0	4

Output: 4 x 4

Pares correspondentes de camadas de downsampling e upsampling

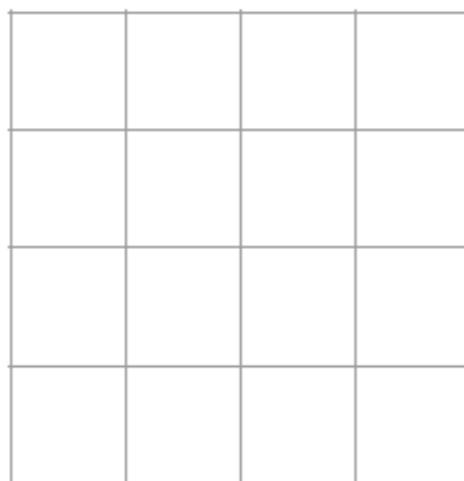


Detecção e segmentação de objetos

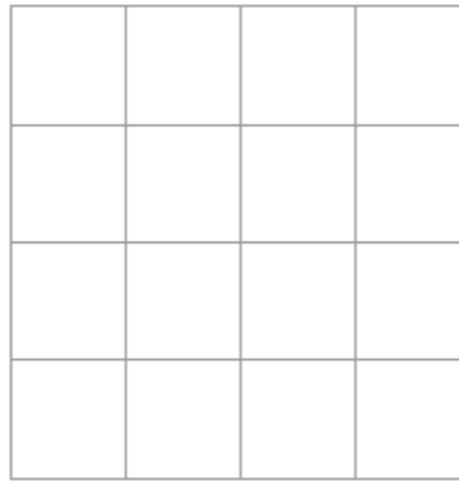
Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

Relembrando: convolução 3x3, stride 1, pad 1



Input: 4 x 4



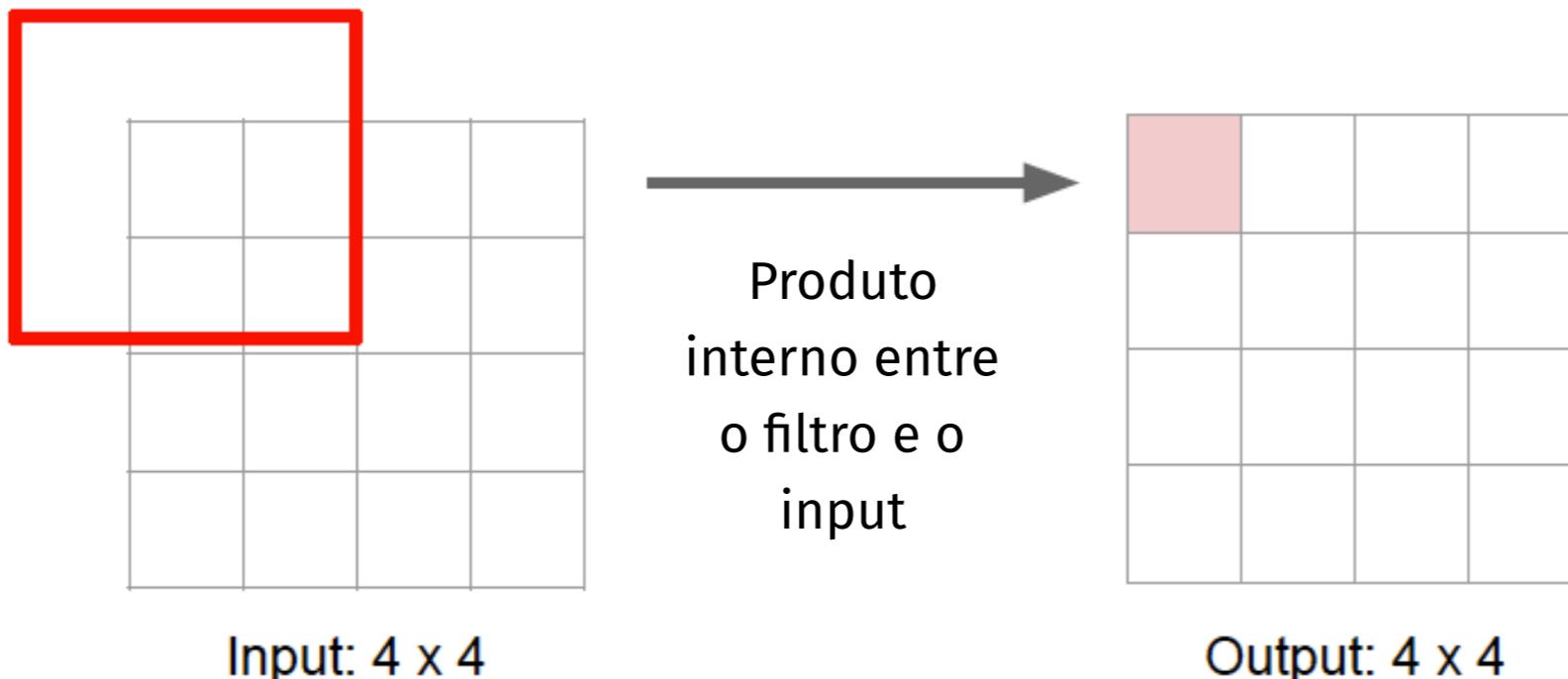
Output: 4 x 4

Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

Relembrando: convolução 3x3, stride 1, pad 1

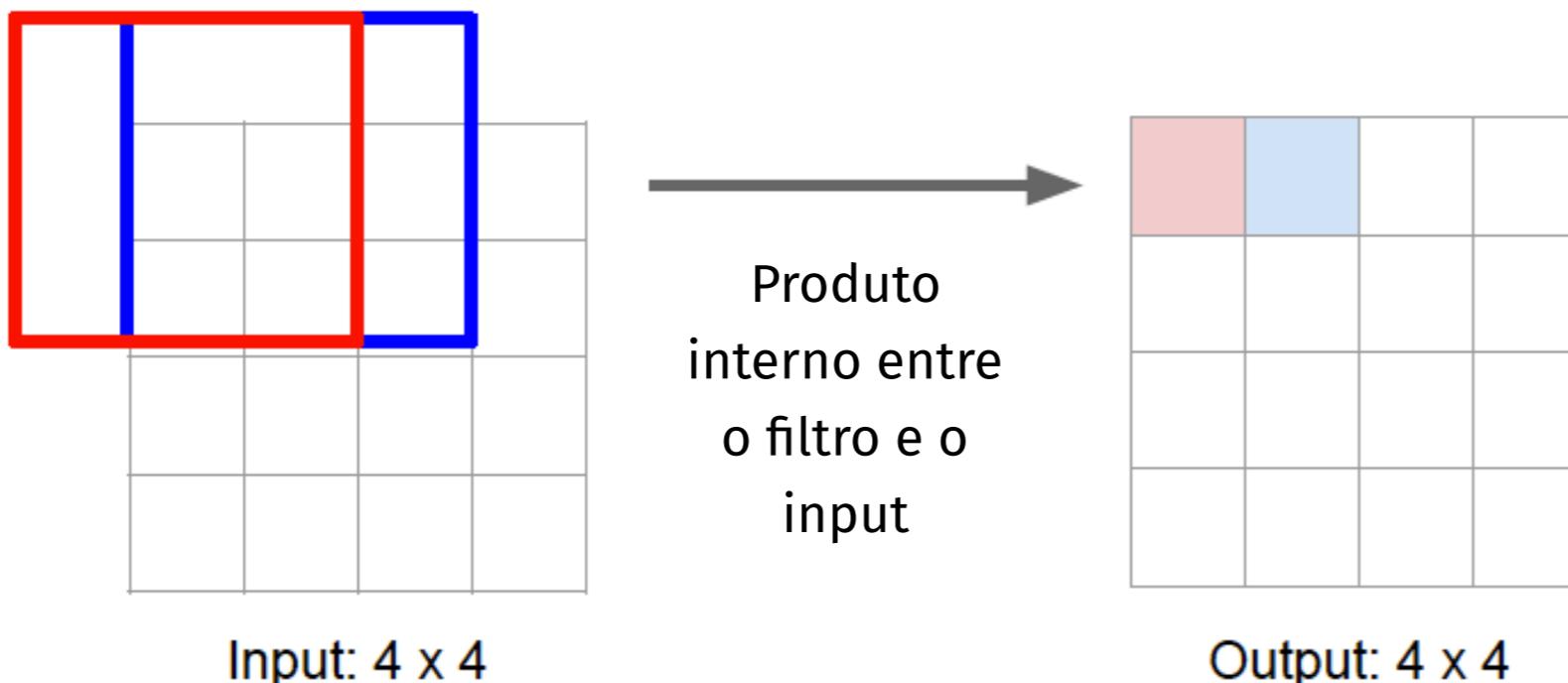


Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

Relembrando: convolução 3x3, stride 1, pad 1

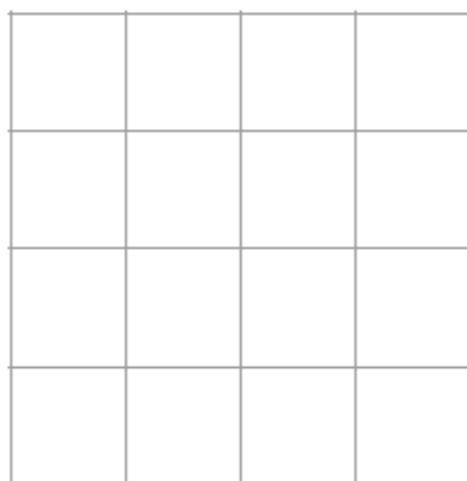


Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

Relembrando: convolução 3x3, stride 2, pad 1



Input: 4 x 4



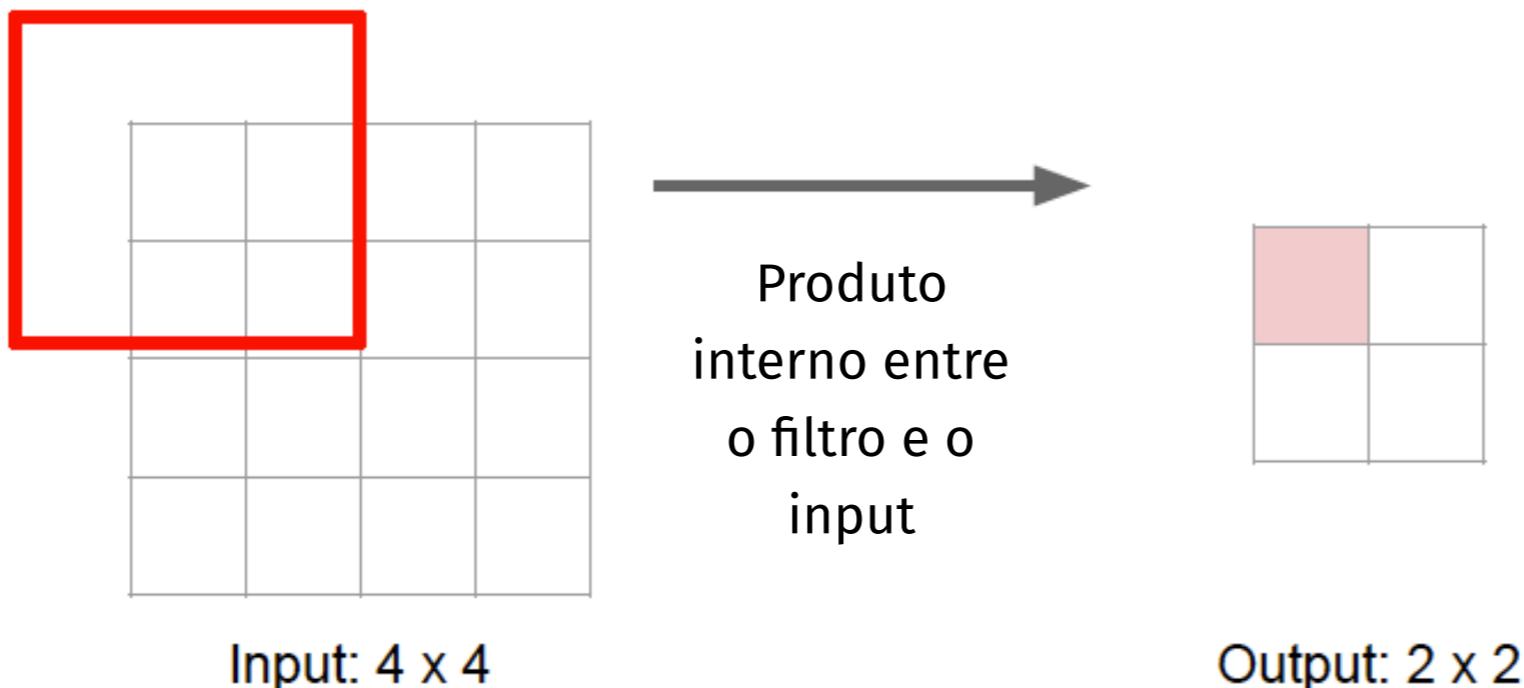
Output: 2 x 2

Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

Relembrando: convolução 3x3, stride 2, pad 1

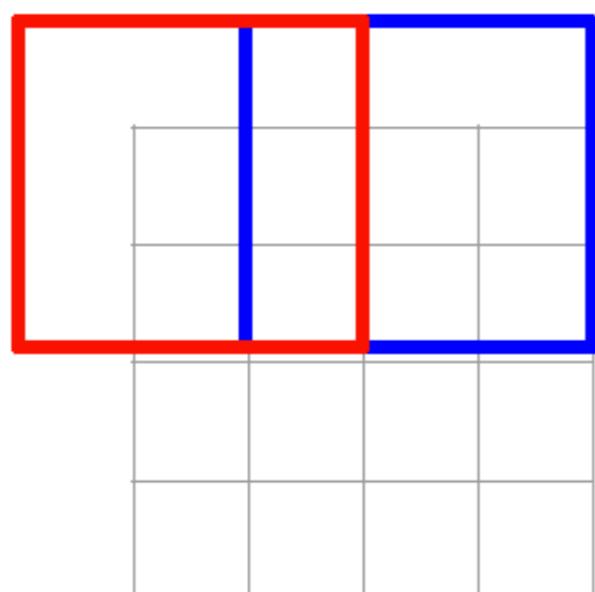


Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

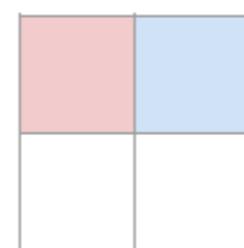
Convolução transporta ou “deconvolution”

Relembrando: convolução 3x3, stride 2, pad 1



Input: 4 x 4

Produto
interno entre
o filtro e o
input



Output: 2 x 2

Filtro move 2 pixels no
input para cada pixel no
output

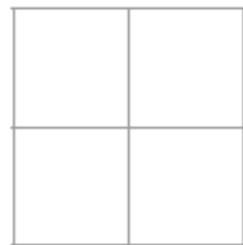
Stride indica a taxa de
movimento entre o input e
o output

Detecção e segmentação de objetos

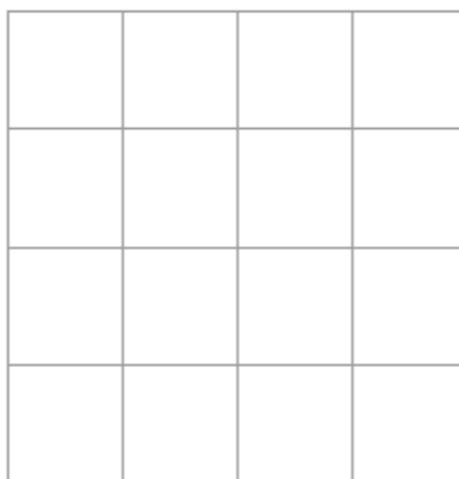
Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

Convolução transposta 3x3, stride 2, pad 1



Input: 2 x 2



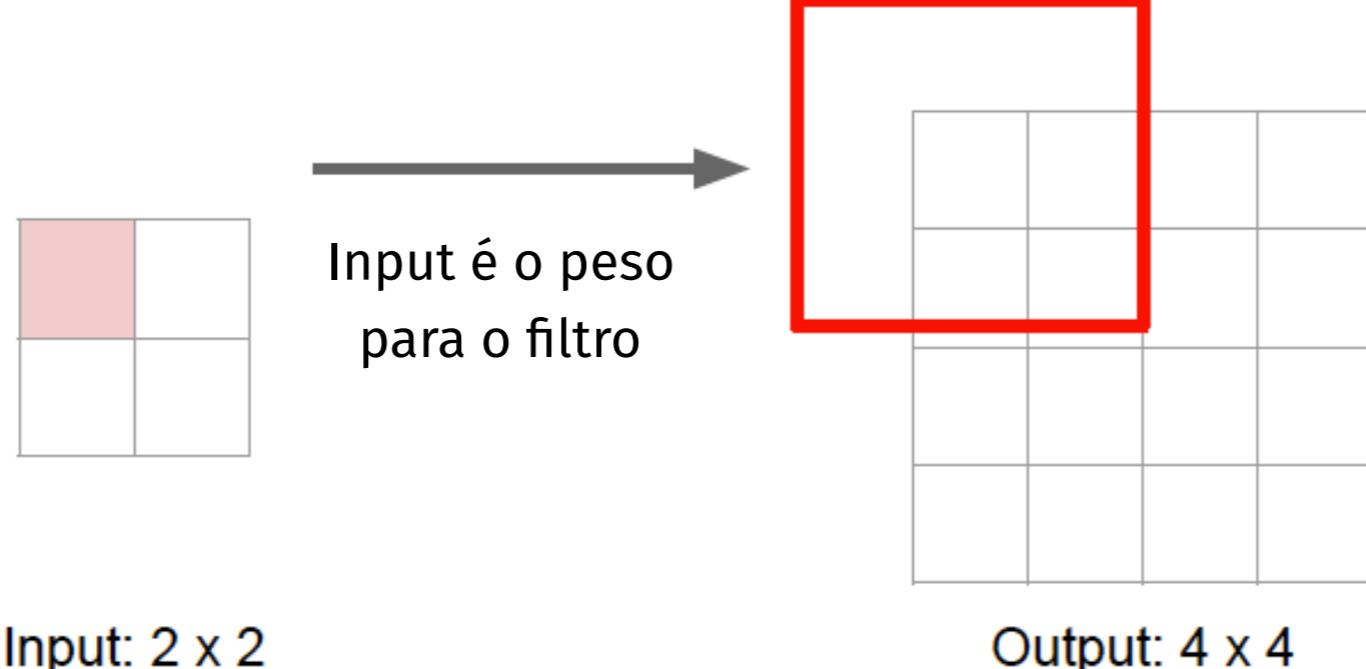
Output: 4 x 4

Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”

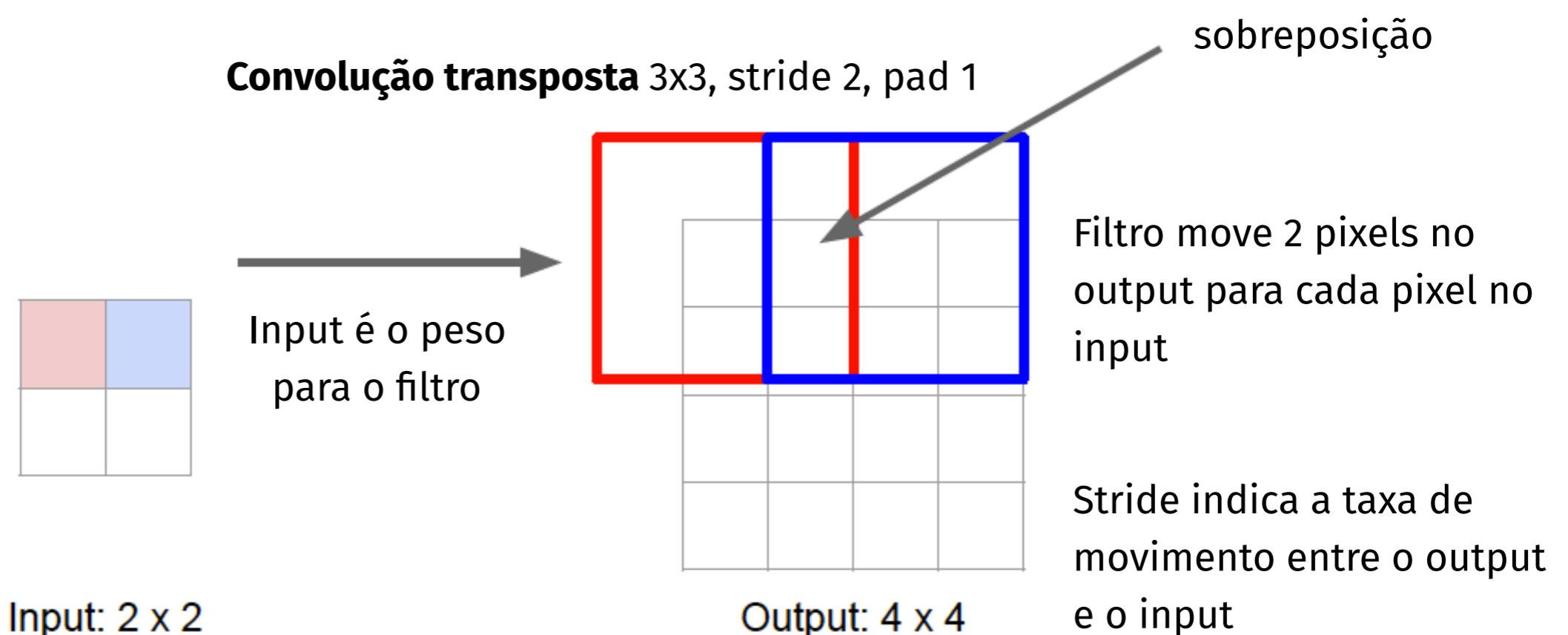
Convolução transposta 3x3, stride 2, pad 1



Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

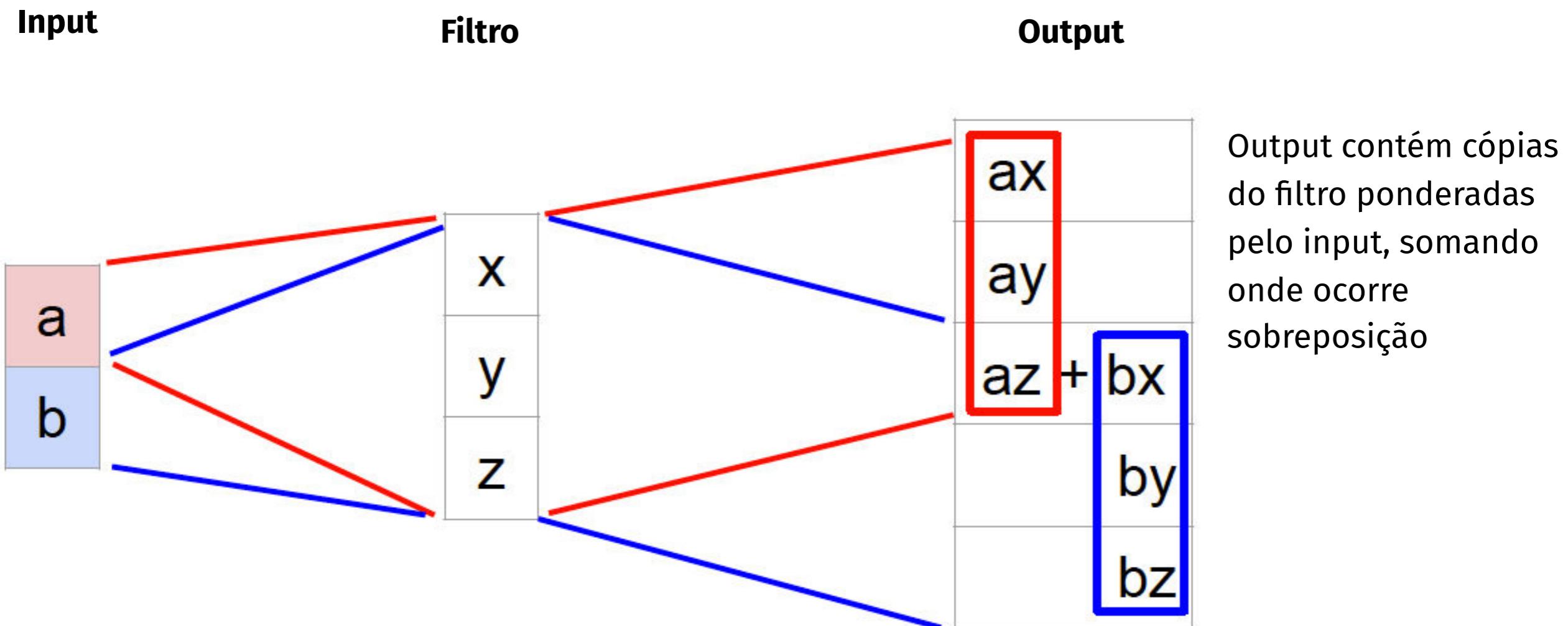
Convolução transporta ou “deconvolution”



Detecção e segmentação de objetos

Upsampling na rede: com aprendizado

Convolução transporta ou “deconvolution”: exemplo 1-D



Detecção e segmentação de objetos

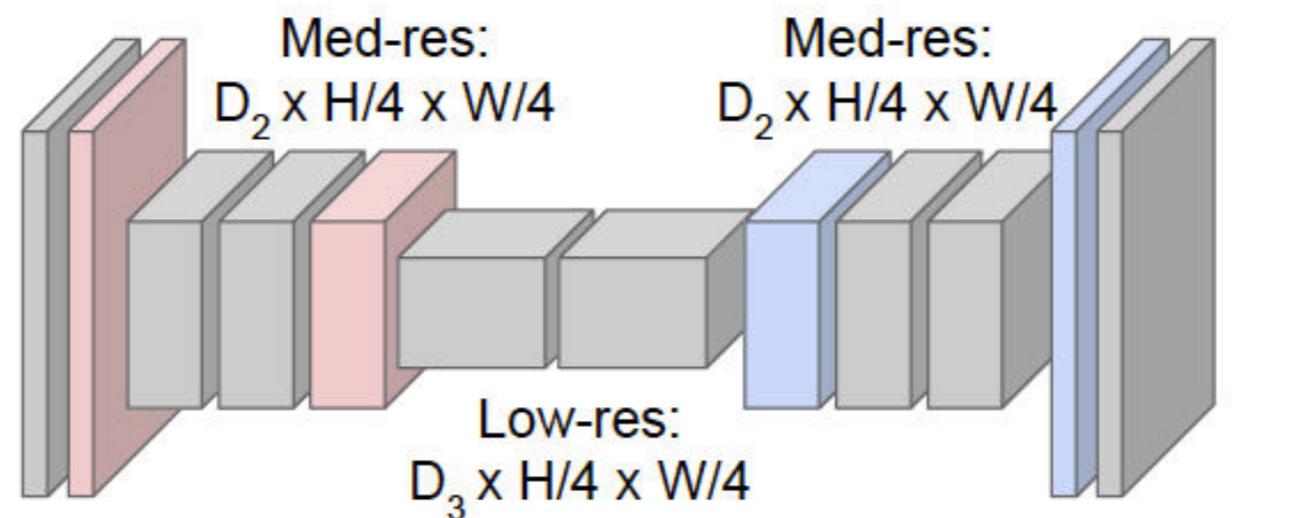
Segmentação semântica: fully convolutional

Downsampling:
pooling,
convolução com
stride



Input:
 $3 \times H \times W$

High-res:
 $D_1 \times H/2 \times W/2$



Projete uma rede com muitas camadas convolucionais,
com **downsampling** e **upsampling** dentro da rede.

Upsampling:
unpooling ou
convolução
transposta com
stride



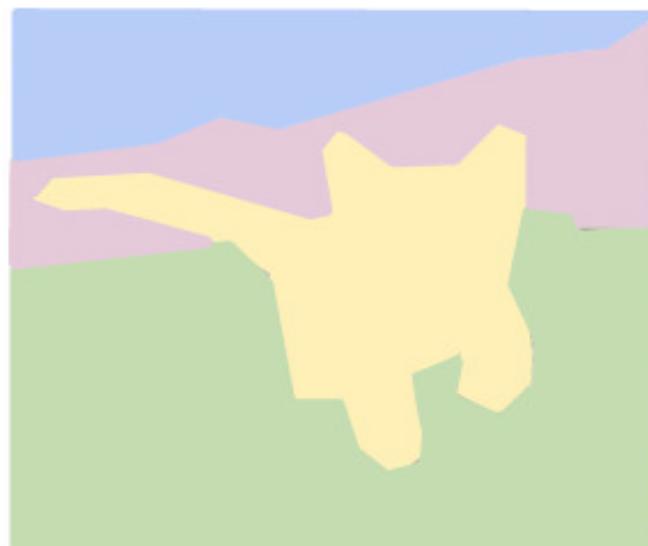
High-res:
 $D_1 \times H/2 \times W/2$

Predictions:
 $H \times W$

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

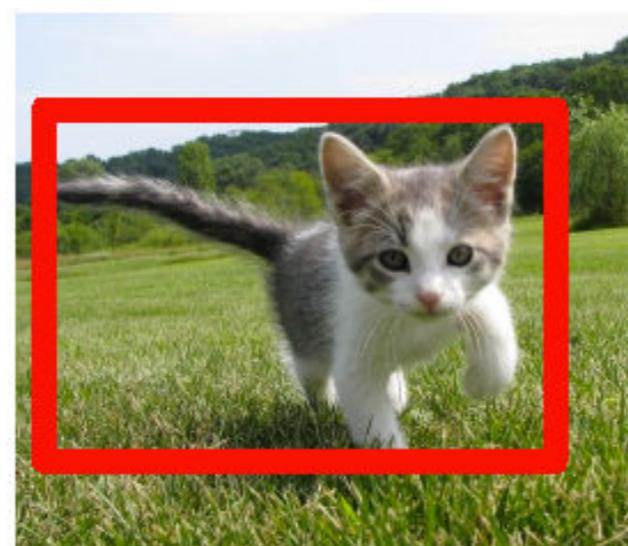
Detecção e segmentação de objetos

Segmentação
semântica
(apenas pixels)



GRASS, CAT,
TREE, SKY

Classificação +
Localização
(objeto único)



CAT

Detecção
(objetos múltiplos)



DOG, DOG, CAT

Segmentação de
instâncias
(objetos múltiplos)



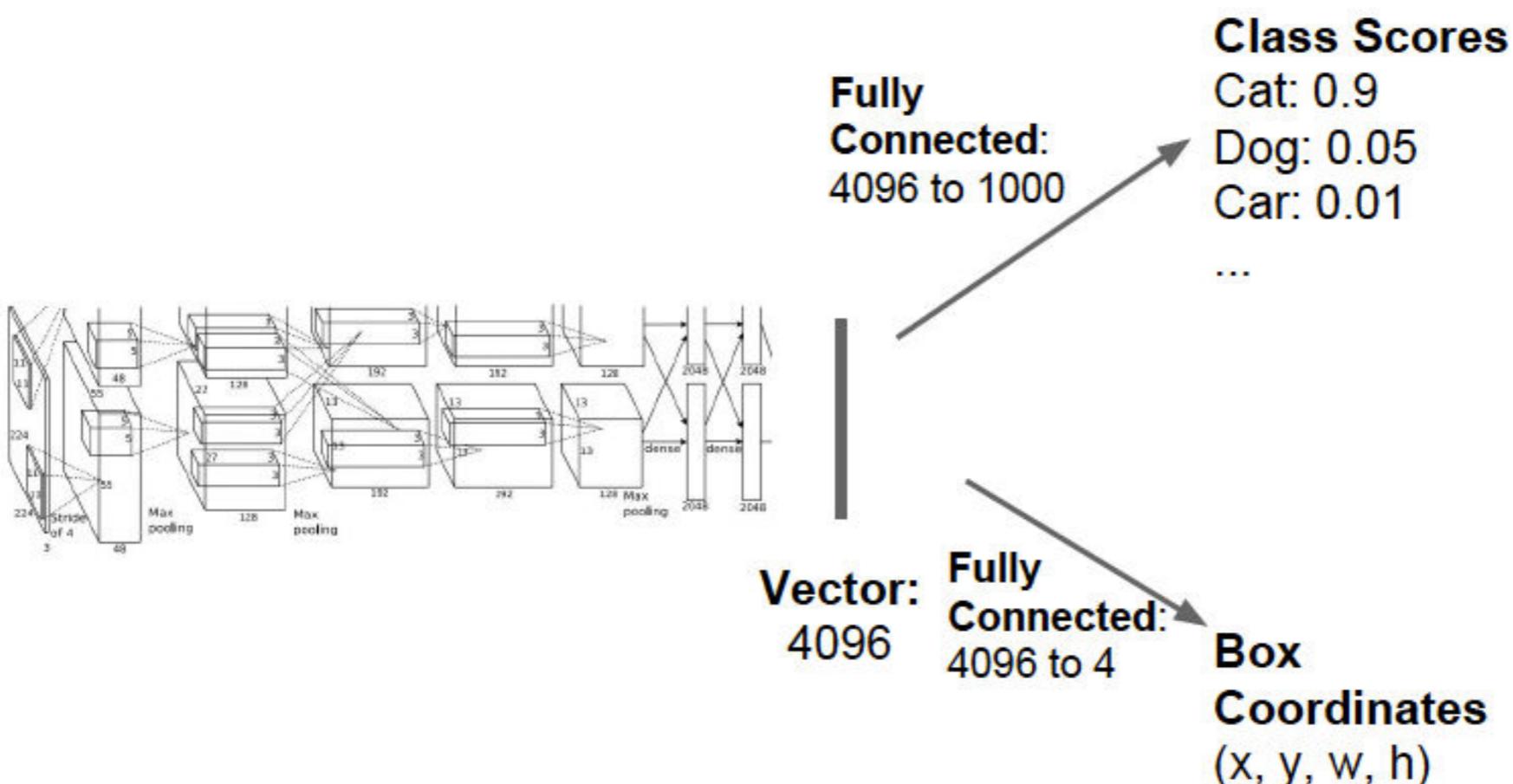
DOG, DOG, CAT

Detecção e segmentação de objetos

Classificação + Localização



This image is CC0 public domain



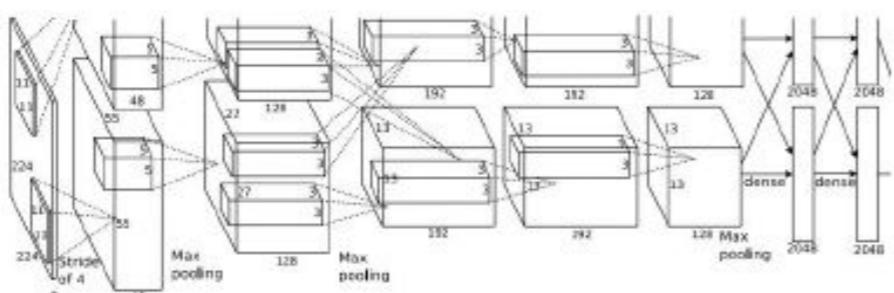
Tratar a localização como um problema de regressão.

Detecção e segmentação de objetos

Classificação + Localização



This image is CC0 public domain



Vector:
4096

Fully
Connected:
4096 to 1000

Class Scores

Cat: 0.9
Dog: 0.05
Car: 0.01
...

Correct label:
Cat

Softmax
Loss

Box
Coordinates → L2 Loss
(x, y, w, h)

Correct box:
(x', y', w', h')

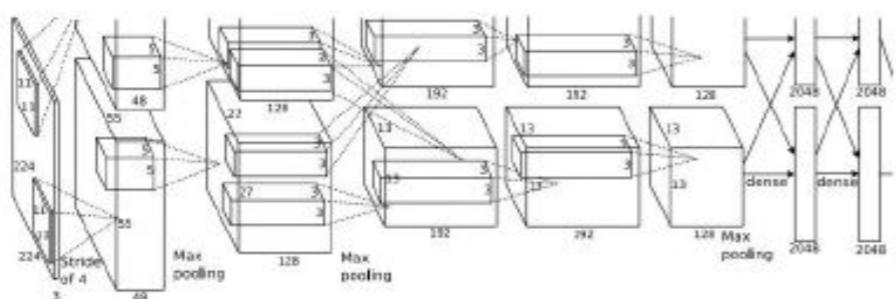
Tratar a localização como um problema de regressão.

Detecção e segmentação de objetos

Classificação + Localização



This image is CC0 public domain



Vector:
4096

Fully
Connected:
4096 to 1000

Class Scores

Cat: 0.9
Dog: 0.05
Car: 0.01
...

Correct label:
Cat

Softmax
Loss

+

Loss

Box

Coordinates → L2 Loss
(x, y, w, h)

Correct box:
(x', y', w', h')

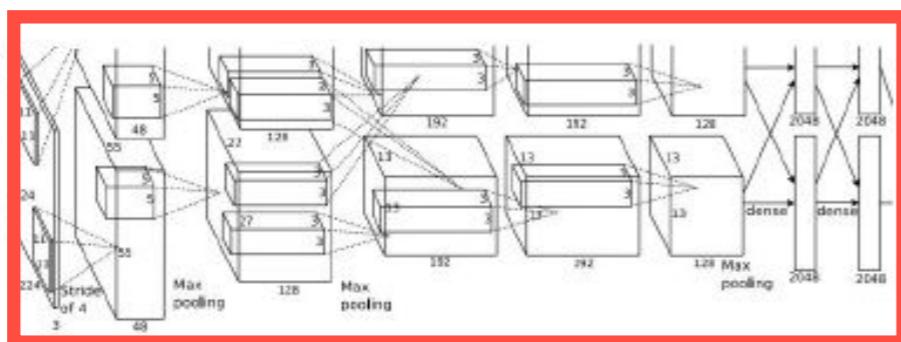
Tratar a localização como um problema de regressão.

Detecção e segmentação de objetos

Classificação + Localização

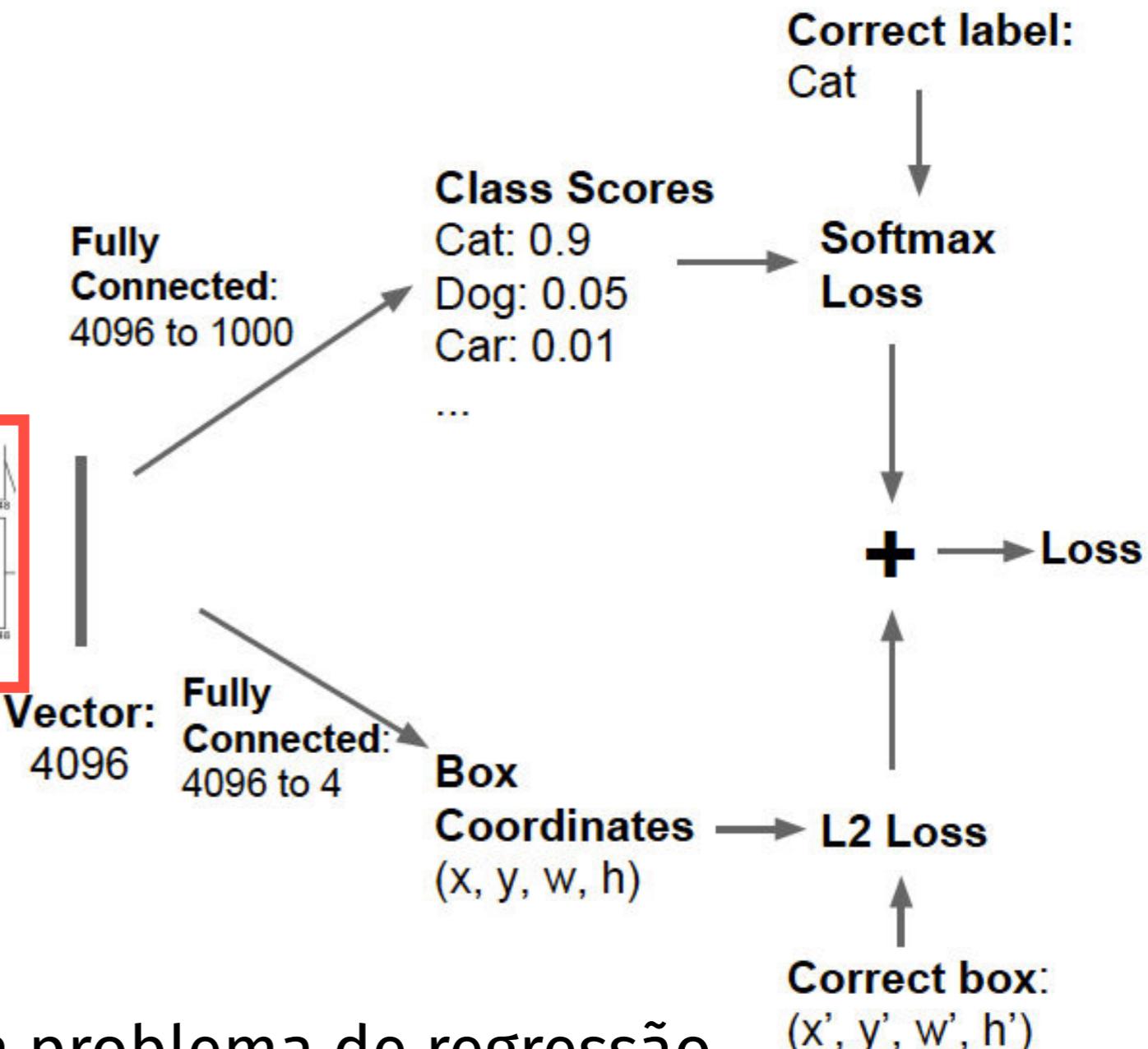


This image is CC0 public domain



Geralmente pré-treinado na ImageNet (transfer learning)

Tratar a localização como um problema de regressão.



Detecção e segmentação de objetos

Classificação + Localização: outros exemplos

Estimativa de pose em humanos



Representa a pose como um conjunto de 14 posições de articulação

- Pés esquerdo/direito
- Joelhos esquerdo/direito
- Quadril esquerdo/direito
- Ombros esquerdo/direito
- Cotovelos esquerdo/direito
- Mãos esquerda/direita
- Pescoço
- Topo da cabeça

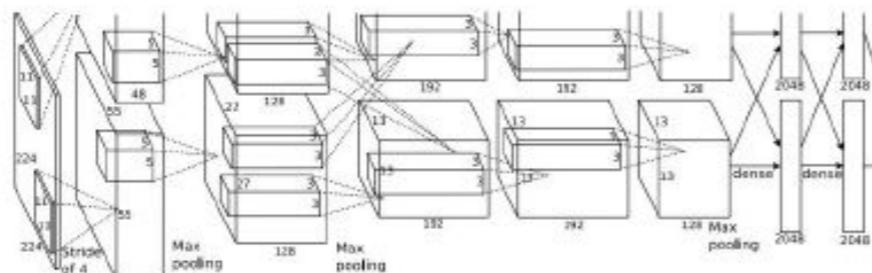
This image is licensed under CC-BY 2.0.

Johnson and Everingham, "Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation", BMVC 2010

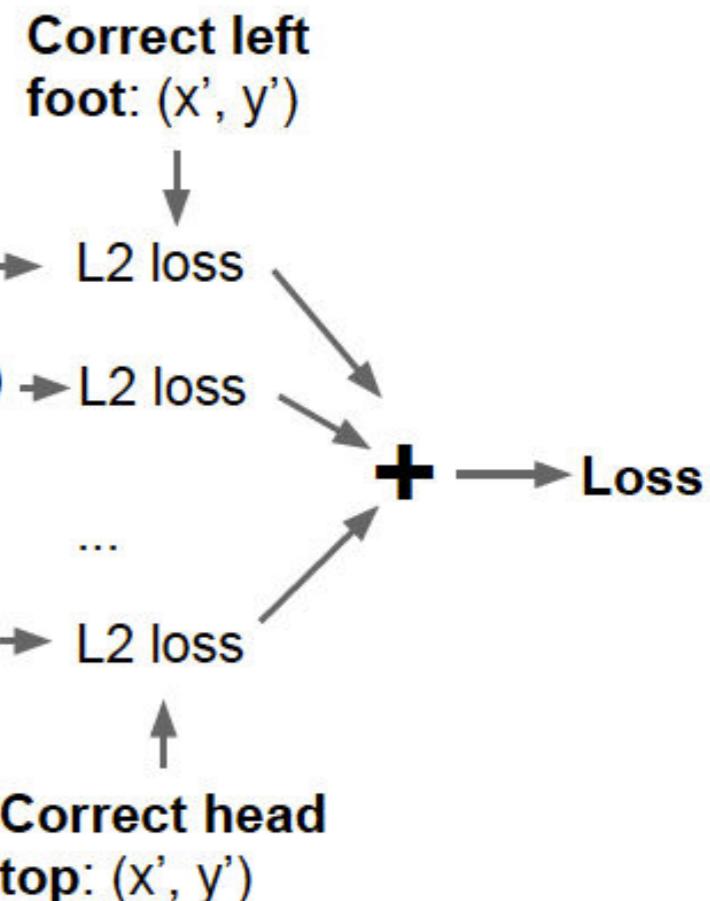
Detecção e segmentação de objetos

Classificação + Localização: outros exemplos

Estimativa de pose em humanos



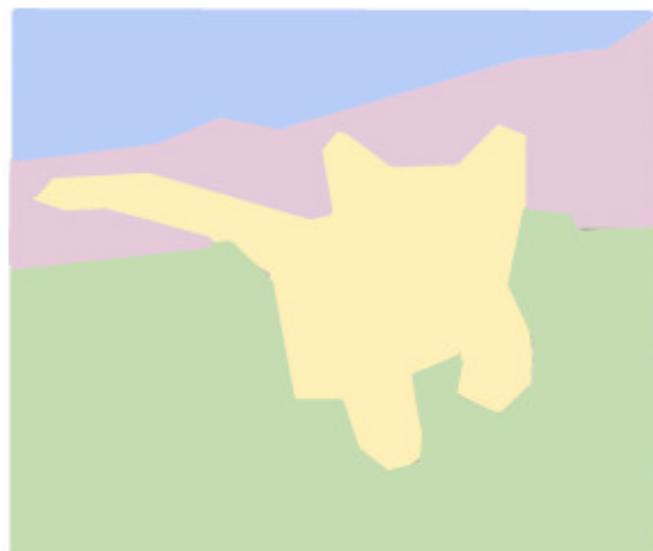
Vector:
4096



Toshev and Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks", CVPR 2014

Detecção e segmentação de objetos

Segmentação
semântica
(apenas pixels)



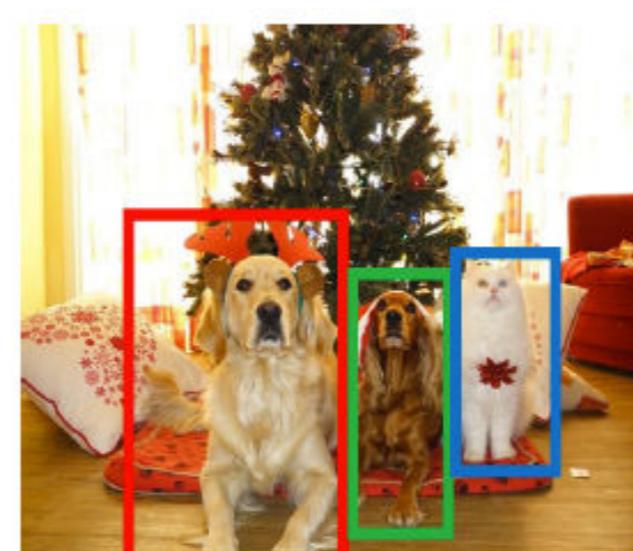
GRASS, CAT,
TREE, SKY

Classificação +
Localização
(objeto único)



CAT

Detecção
(objetos múltiplos)



DOG, DOG, CAT

Segmentação de
instâncias
(objetos múltiplos)



DOG, DOG, CAT

Detecção e segmentação de objetos

Detecção de objetos

Impacto do deep learning

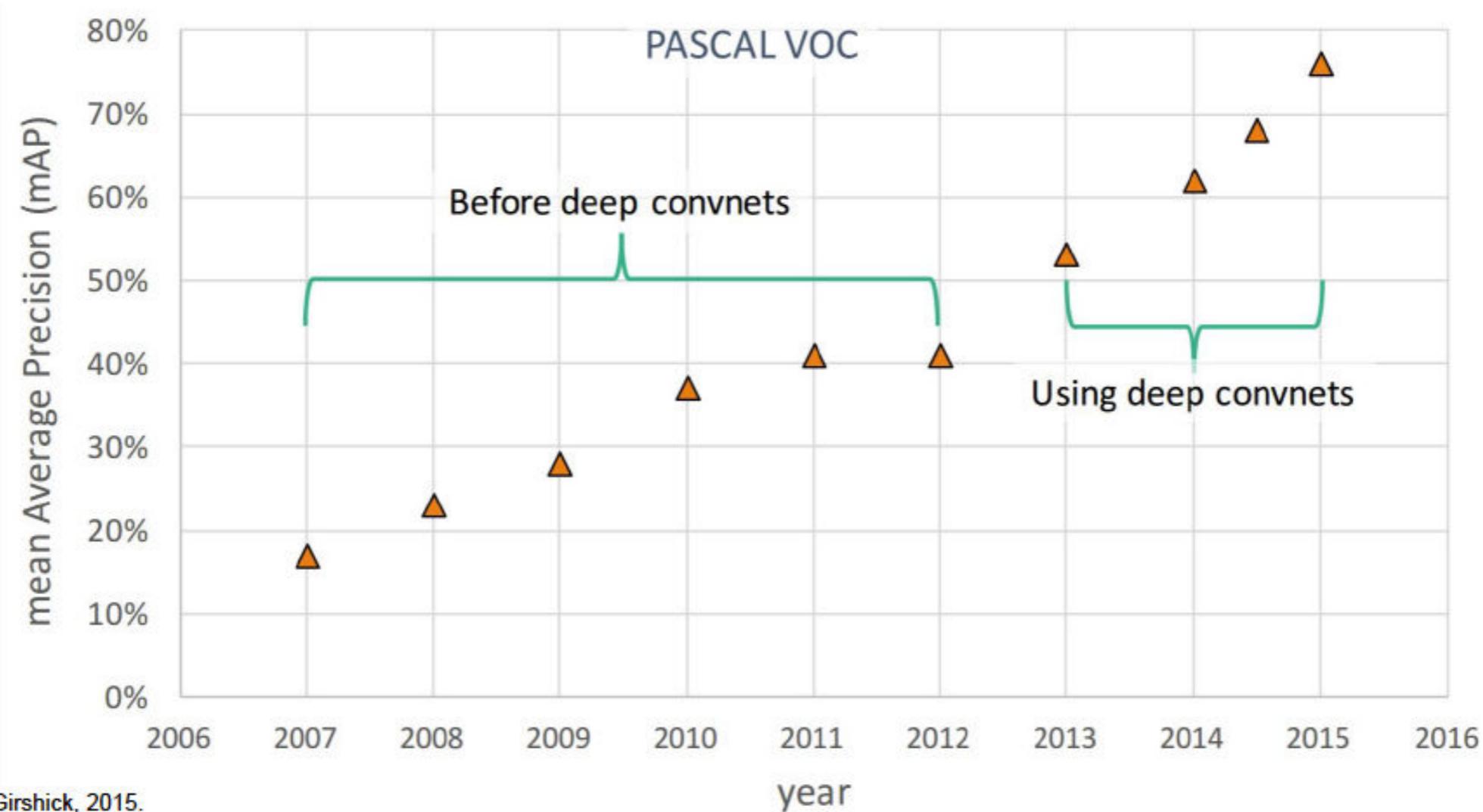
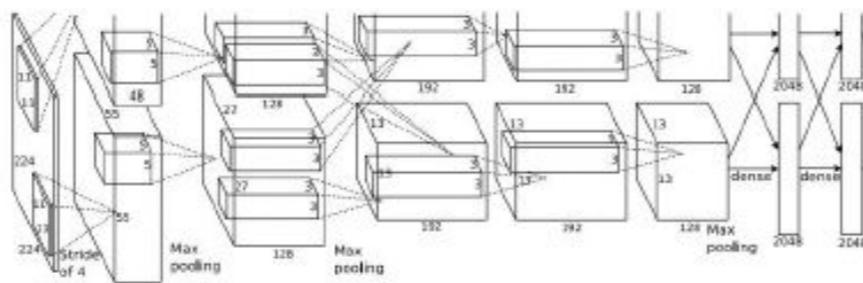


Figure copyright Ross Girshick, 2015.
Reproduced with permission.

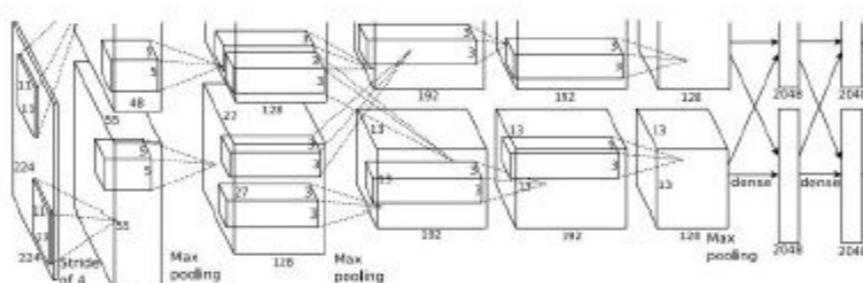
Detecção e segmentação de objetos

Detecção de objetos

Régressão?



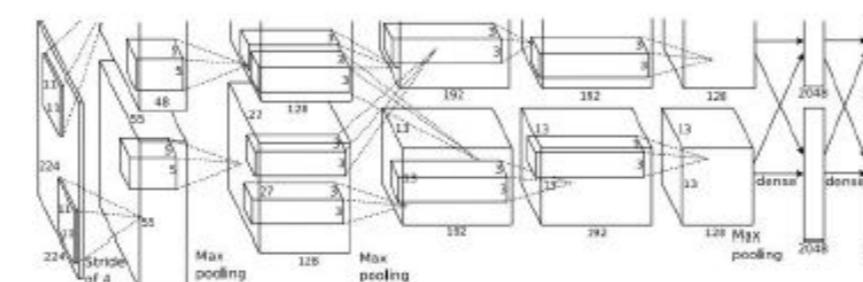
CAT: (x, y, w, h)



DOG: (x, y, w, h)

DOG: (x, y, w, h)

CAT: (x, y, w, h)



DUCK: (x, y, w, h)

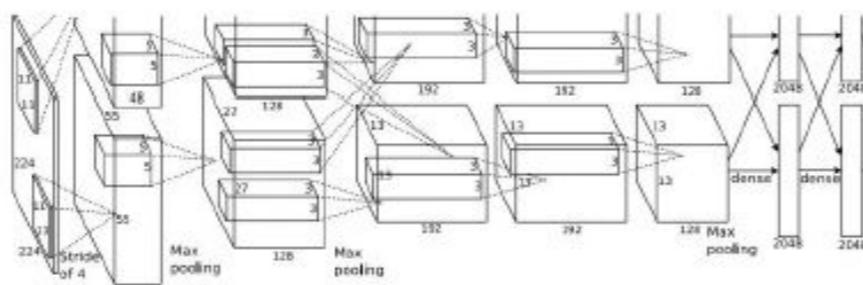
DUCK: (x, y, w, h)

....

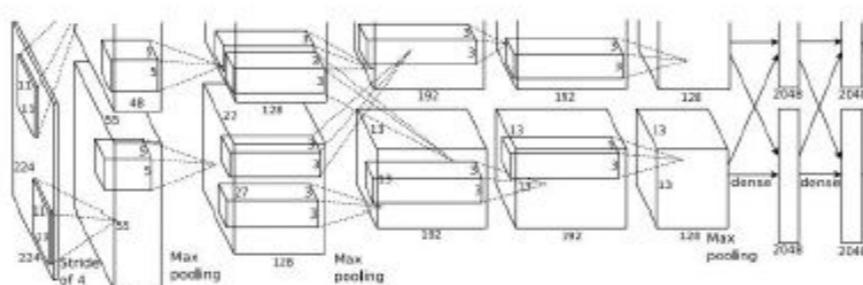
Detecção e segmentação de objetos

Detecção de objetos

Regressão? **Problema!** Cada imagem precisa de uma quantidade diferente de outputs.



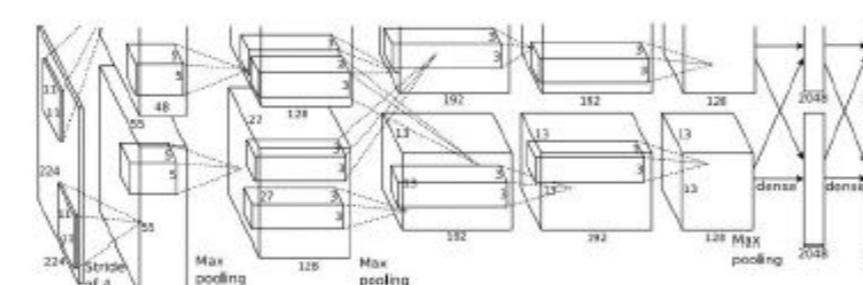
CAT: (x, y, w, h)



DOG: (x, y, w, h)

DOG: (x, y, w, h)

CAT: (x, y, w, h)



DUCK: (x, y, w, h)

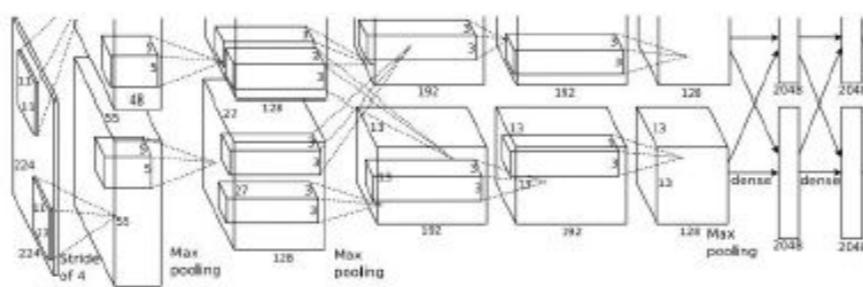
DUCK: (x, y, w, h)

....

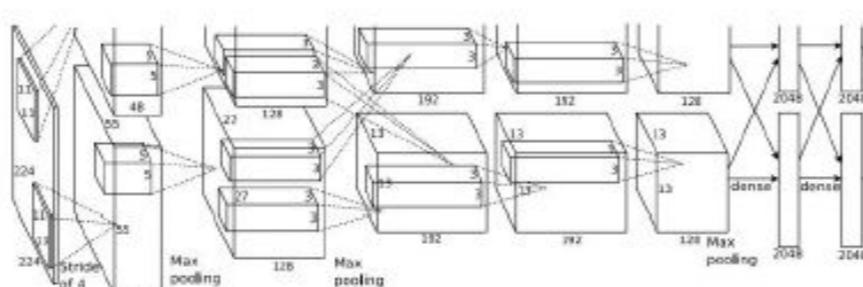
Detecção e segmentação de objetos

Detecção de objetos

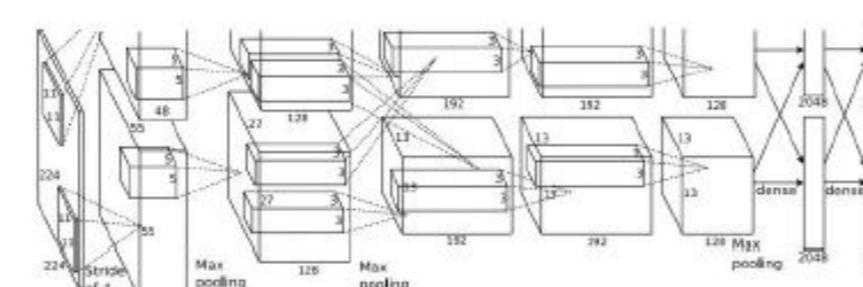
Regressão? **Problema!** Cada imagem precisa de uma quantidade diferente de outputs.



CAT: (x, y, w, h) 4 números



DOG: (x, y, w, h)
DOG: (x, y, w, h)
CAT: (x, y, w, h)

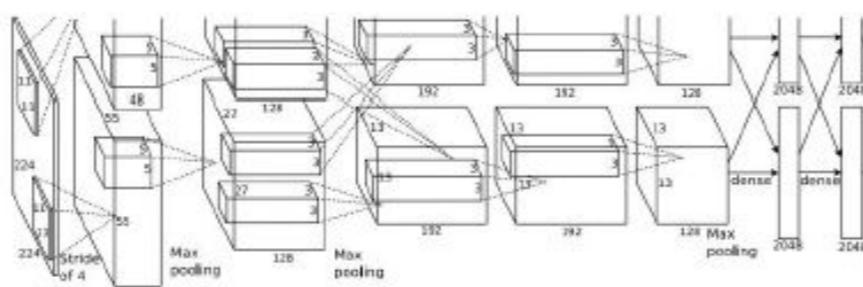


DUCK: (x, y, w, h)
DUCK: (x, y, w, h)
....

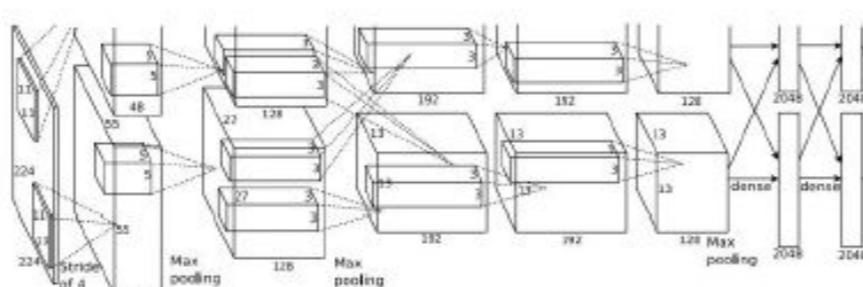
Detecção e segmentação de objetos

Detecção de objetos

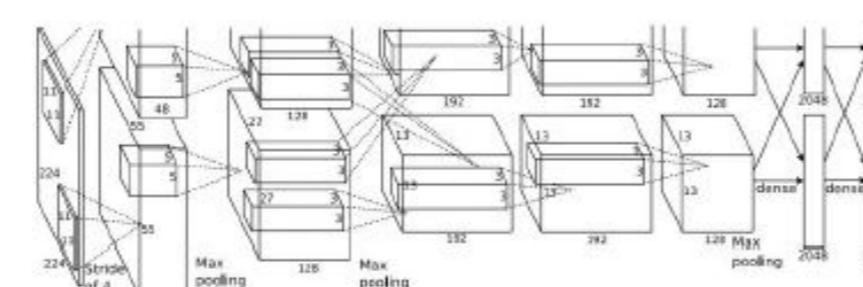
Regressão? **Problema!** Cada imagem precisa de uma quantidade diferente de outputs.



CAT: (x, y, w, h) 4 números



DOG: (x, y, w, h) 16 números
DOG: (x, y, w, h)
CAT: (x, y, w, h)

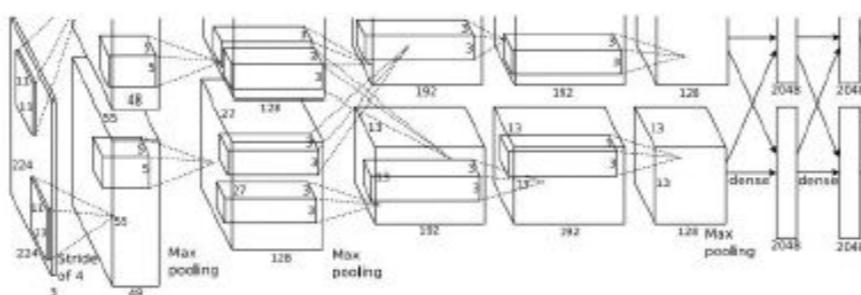


DUCK: (x, y, w, h)
DUCK: (x, y, w, h)
....

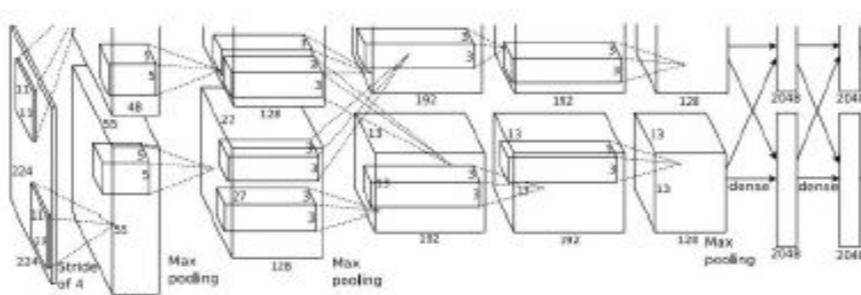
Detecção e segmentação de objetos

Detecção de objetos

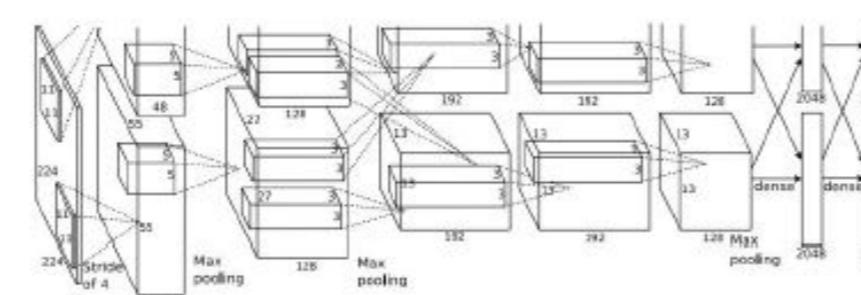
Regressão? **Problema!** Cada imagem precisa de uma quantidade diferente de outputs.



CAT: (x, y, w, h) 4 números



DOG: (x, y, w, h) 16 números
DOG: (x, y, w, h)
CAT: (x, y, w, h)

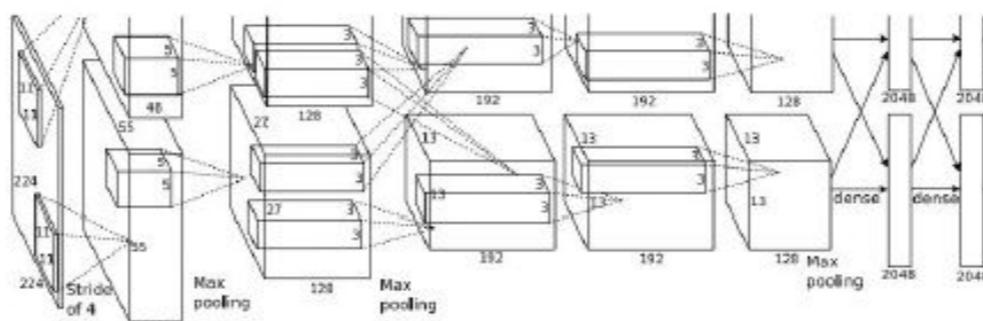


DUCK: (x, y, w, h) Muitos números
DUCK: (x, y, w, h)
....

Detecção e segmentação de objetos

Detecção de objetos

Classificação: janela deslizante

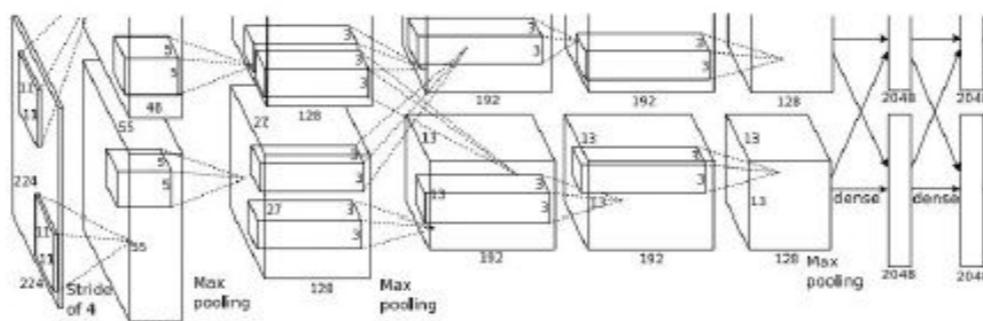


DOG	NO
CAT	NO
BACKGROUND	YES

Detecção e segmentação de objetos

Detecção de objetos

Classificação: janela deslizante



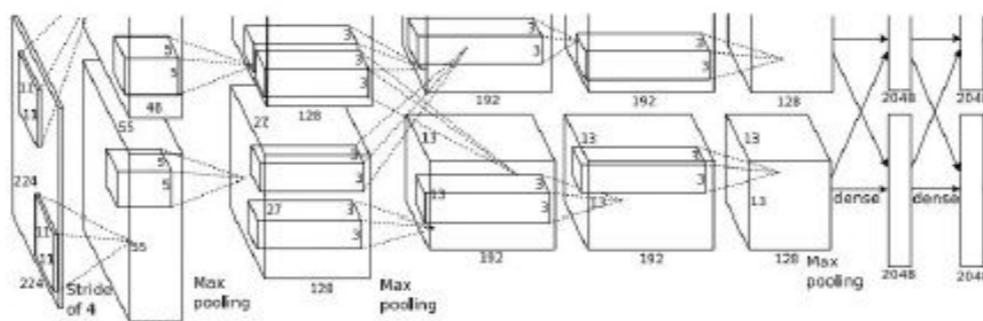
DOG
CAT
BACKGROUND

YES
NO
NO

Detecção e segmentação de objetos

Detecção de objetos

Classificação: janela deslizante

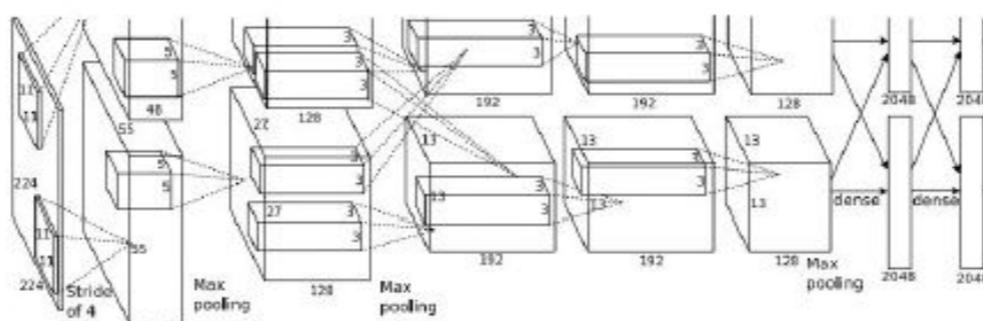


DOG	YES
CAT	NO
BACKGROUND	NO

Detecção e segmentação de objetos

Detecção de objetos

Classificação: janela deslizante



DOG

CAT

BACKGROUND

NO

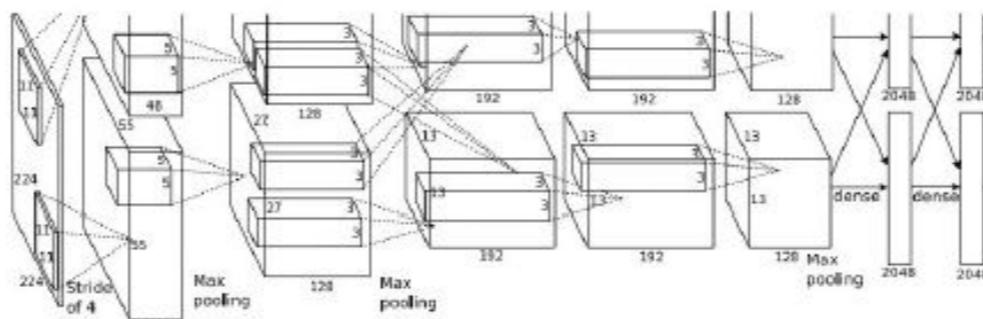
YES

NO

Detecção e segmentação de objetos

Detecção de objetos

Classificação: janela deslizante



DOG	NO
CAT	YES
BACKGROUND	NO

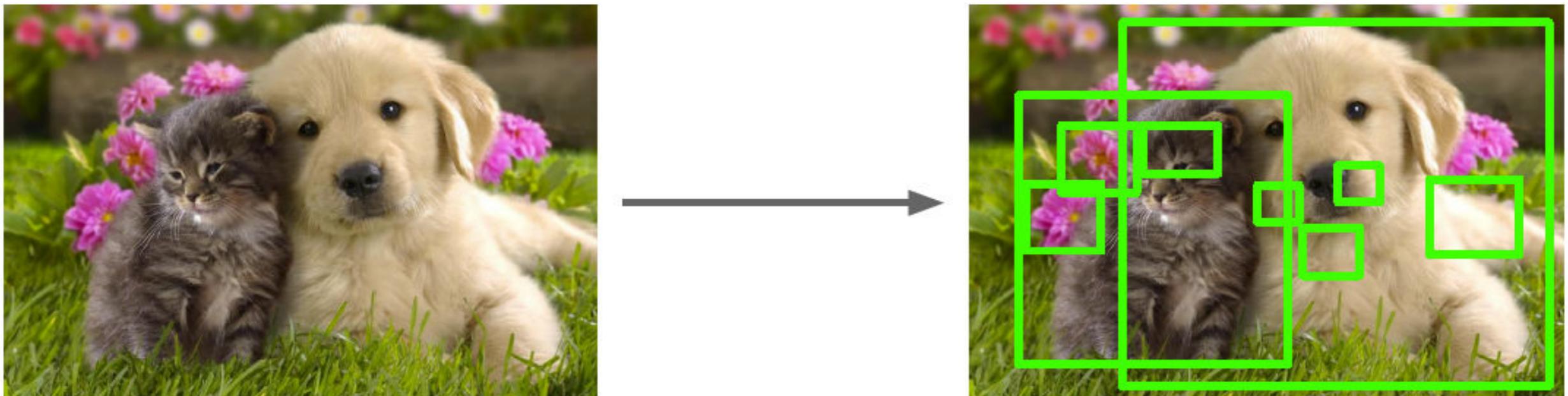
Problema! É preciso aplicar a CNN para um número alto de localidades e escalas, o que é computacionalmente caro!

Detecção e segmentação de objetos

Detecção de objetos

Proposta de regiões

- Achar regiões da imagem onde é mais provável conter objetos
- Relativamente rápido: Selective Search retorna 2000 propostas de regiões em poucos segundos de CPU.



Alexe et al, "Measuring the objectness of image windows", TPAMI 2012

Uijlings et al, "Selective Search for Object Recognition", IJCV 2013

Cheng et al, "BING: Binarized normed gradients for objectness estimation at 300fps", CVPR 2014

Zitnick and Dollar, "Edge boxes: Locating object proposals from edges", ECCV 2014

Detecção e segmentação de objetos

Detecção de objetos: R-CNN



Input image

Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: R-CNN

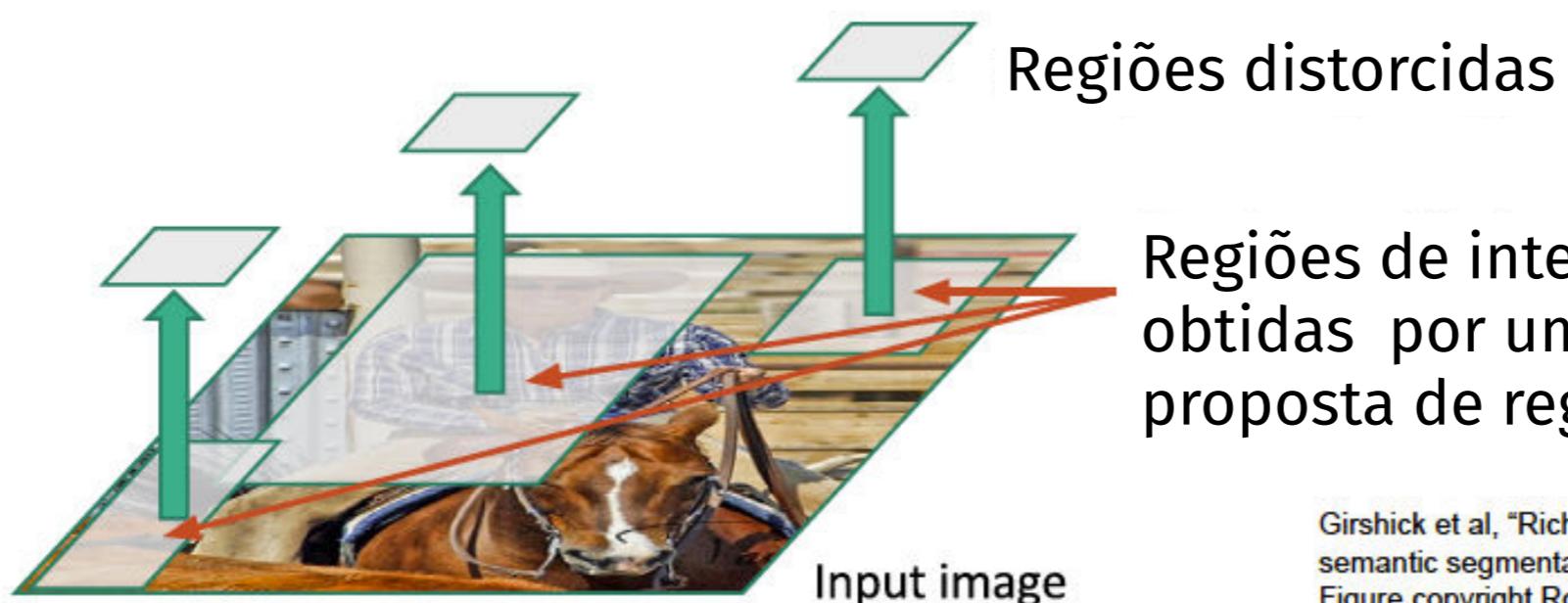


Regiões de interesse (RoI)
obtidas por um método de
proposta de regiões (~2k)

Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: R-CNN

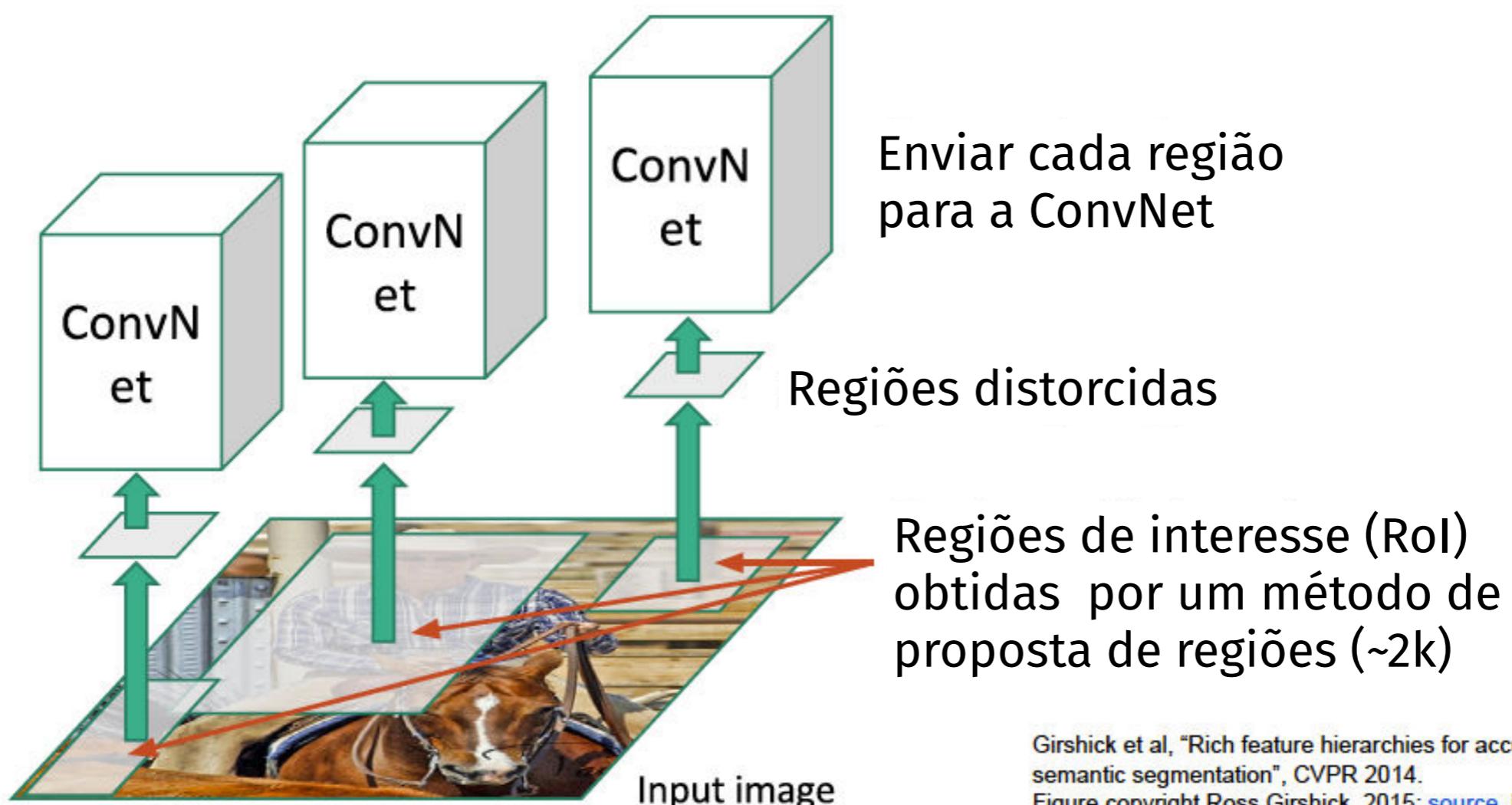


Regiões de interesse (RoI)
obtidas por um método de
proposta de regiões (~2k)

Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

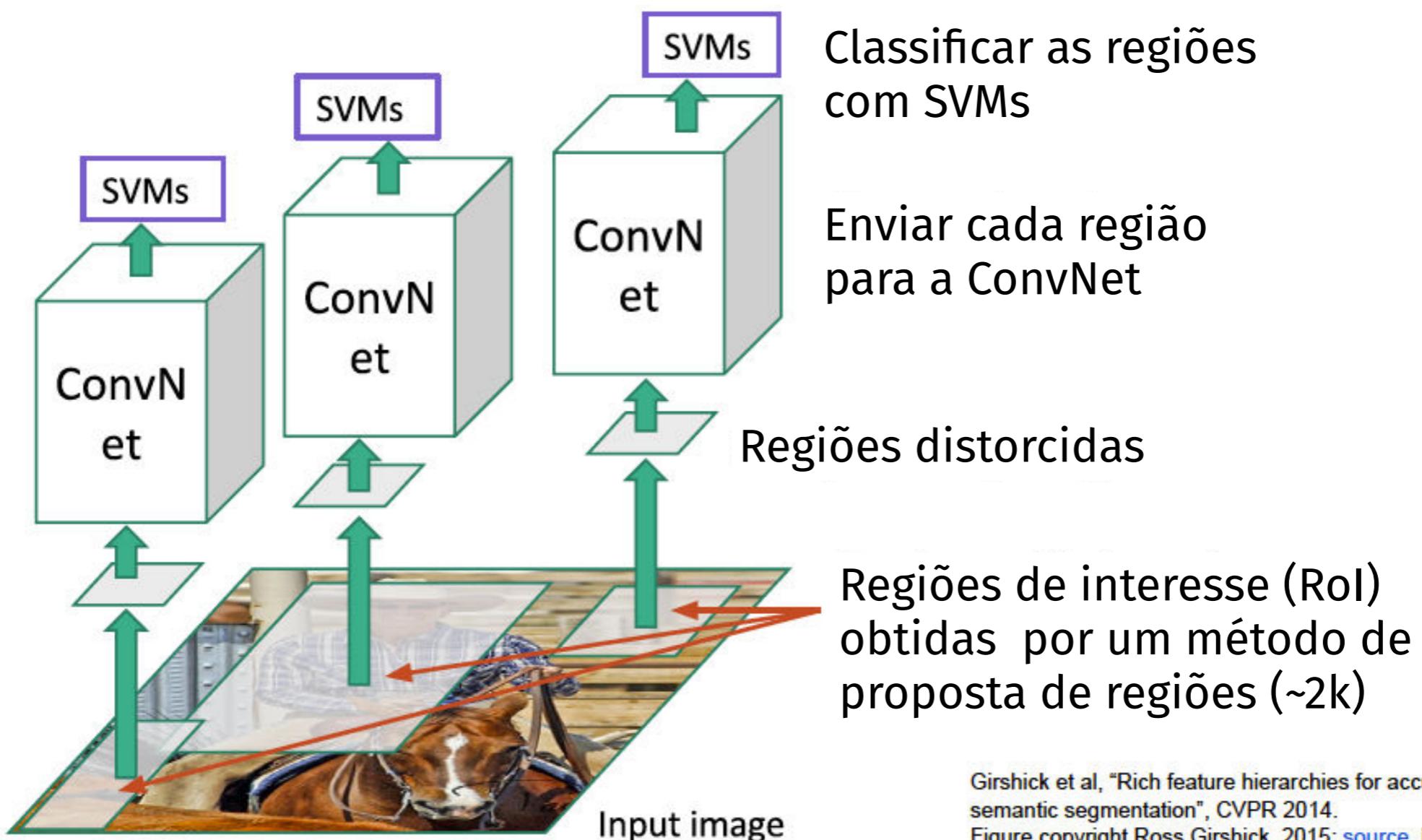
Detecção de objetos: R-CNN



Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

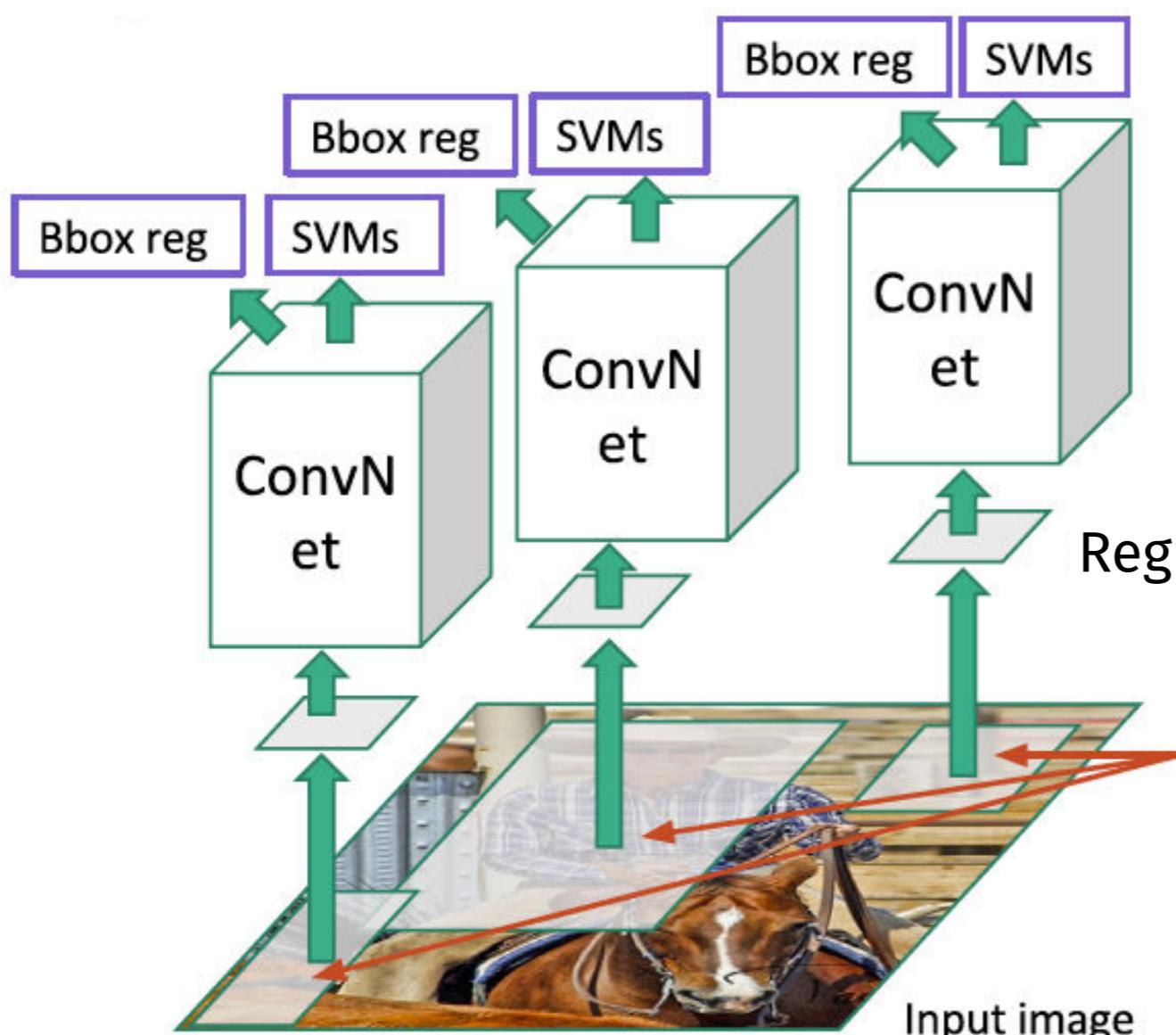
Detecção de objetos: R-CNN



Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: R-CNN



Regressão linear para encontrar os offsets dos bounding boxes

Classificar as regiões com SVMs

Enviar cada região para a ConvNet

Regiões distorcidas

Regiões de interesse (RoI)
obtidas por um método de
proposta de regiões (~2k)

Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: R-CNN

Problemas!

- Objetivos do treinamento:
 - Log loss: fine-tuning da rede com softmax
 - Hinge loss: treinamento dos SVMs
 - Mínimos quadrados: treinamento da regressão dos bounding boxes
- Treinamento é lento (~84h) e ocupa muito espaço em disco
- Inferência (detecção) é lenta: 47 s./imagem com a VGG16

Detecção e segmentação de objetos

Detecção de objetos: Fast R-CNN

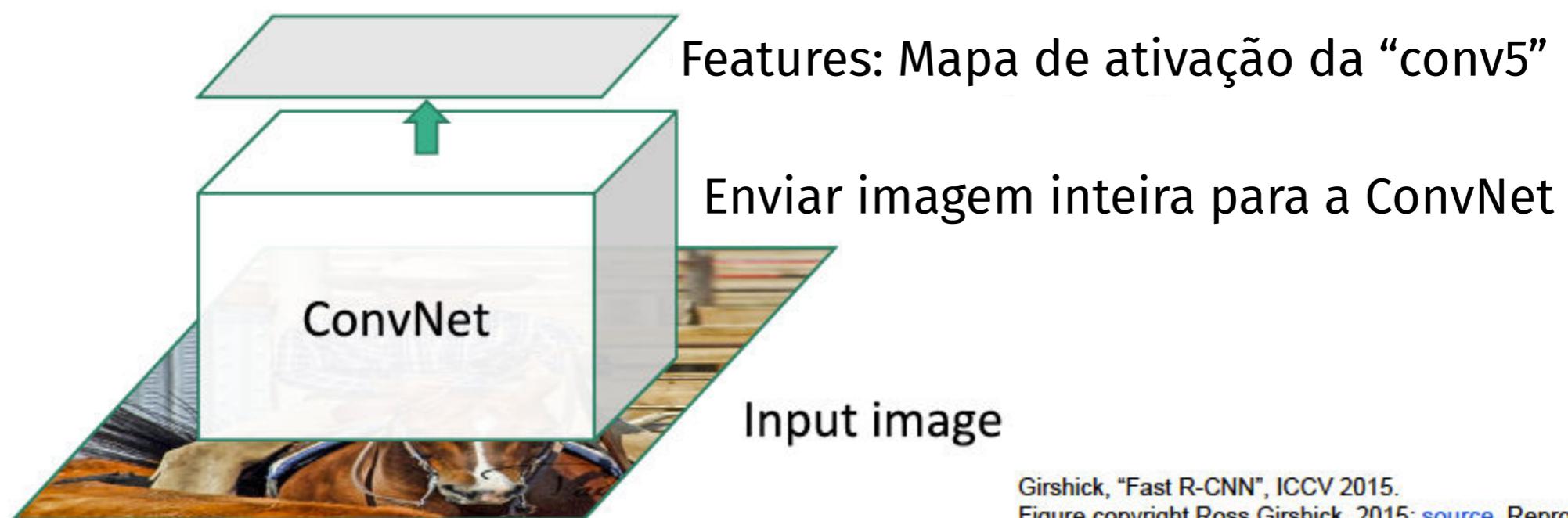


Input image

Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

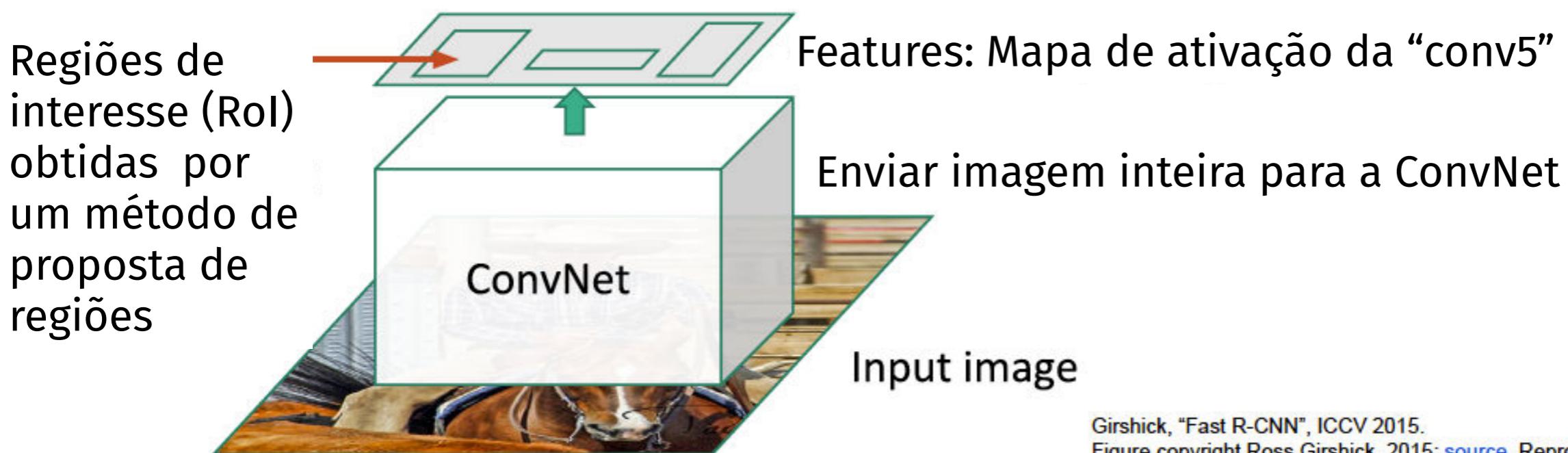
Detecção de objetos: Fast R-CNN



Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

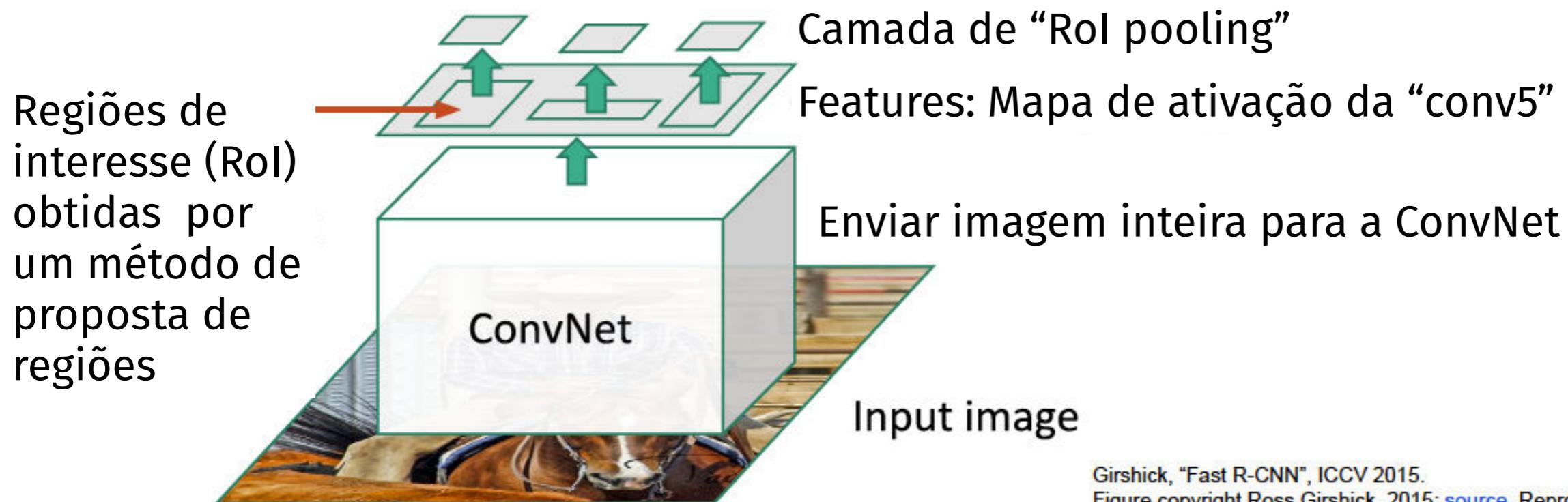
Detecção de objetos: Fast R-CNN



Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

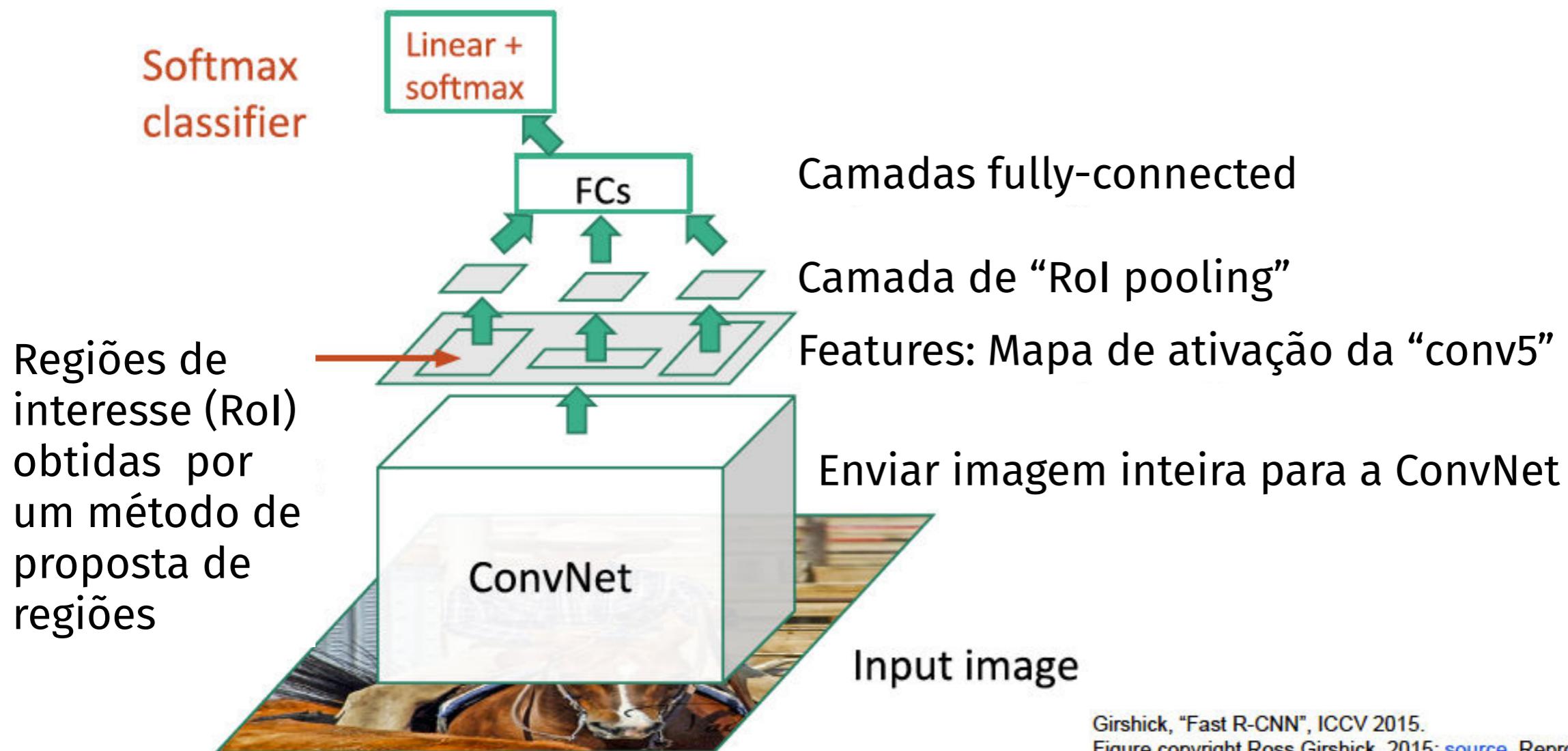
Detecção de objetos: Fast R-CNN



Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

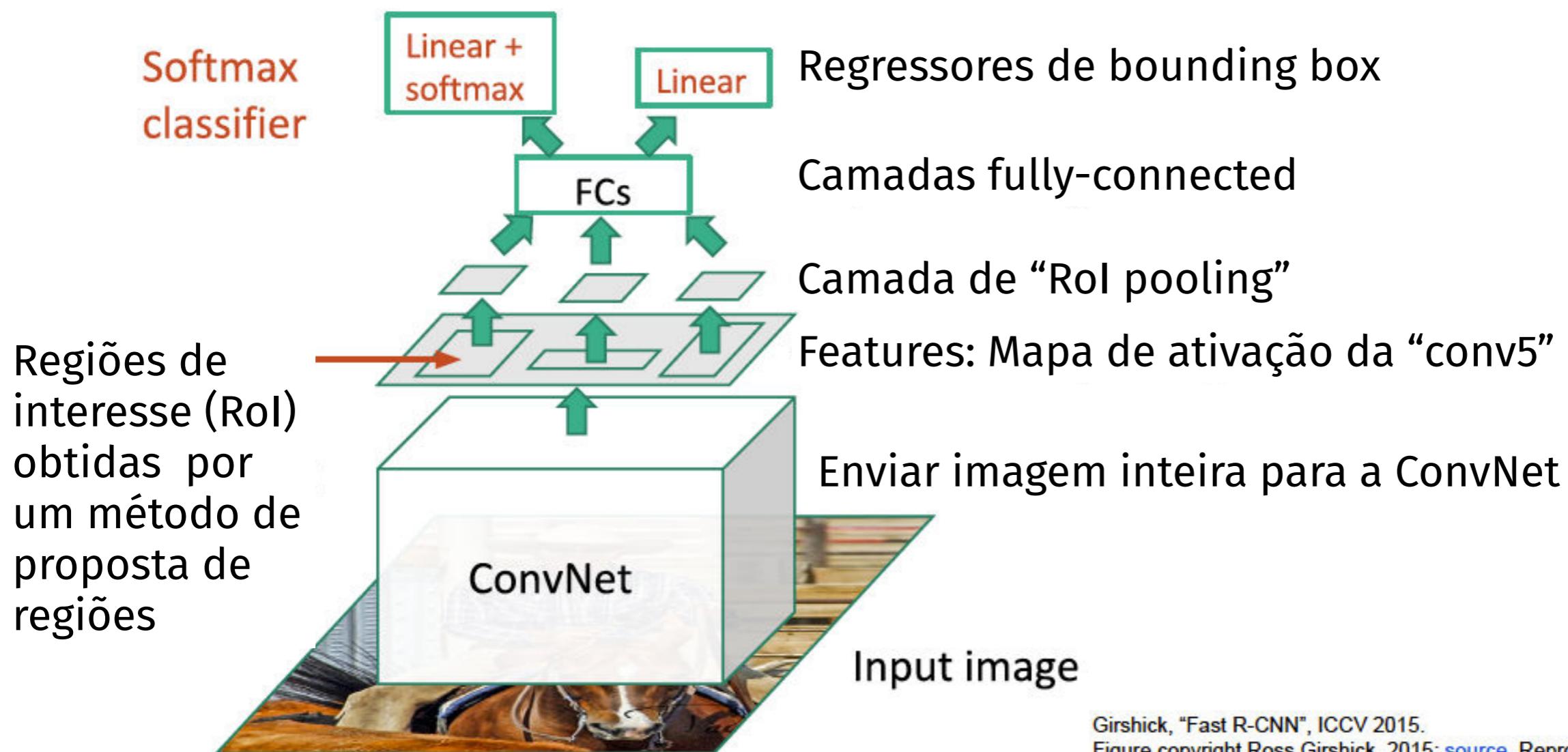
Detecção de objetos: Fast R-CNN



Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: Fast R-CNN

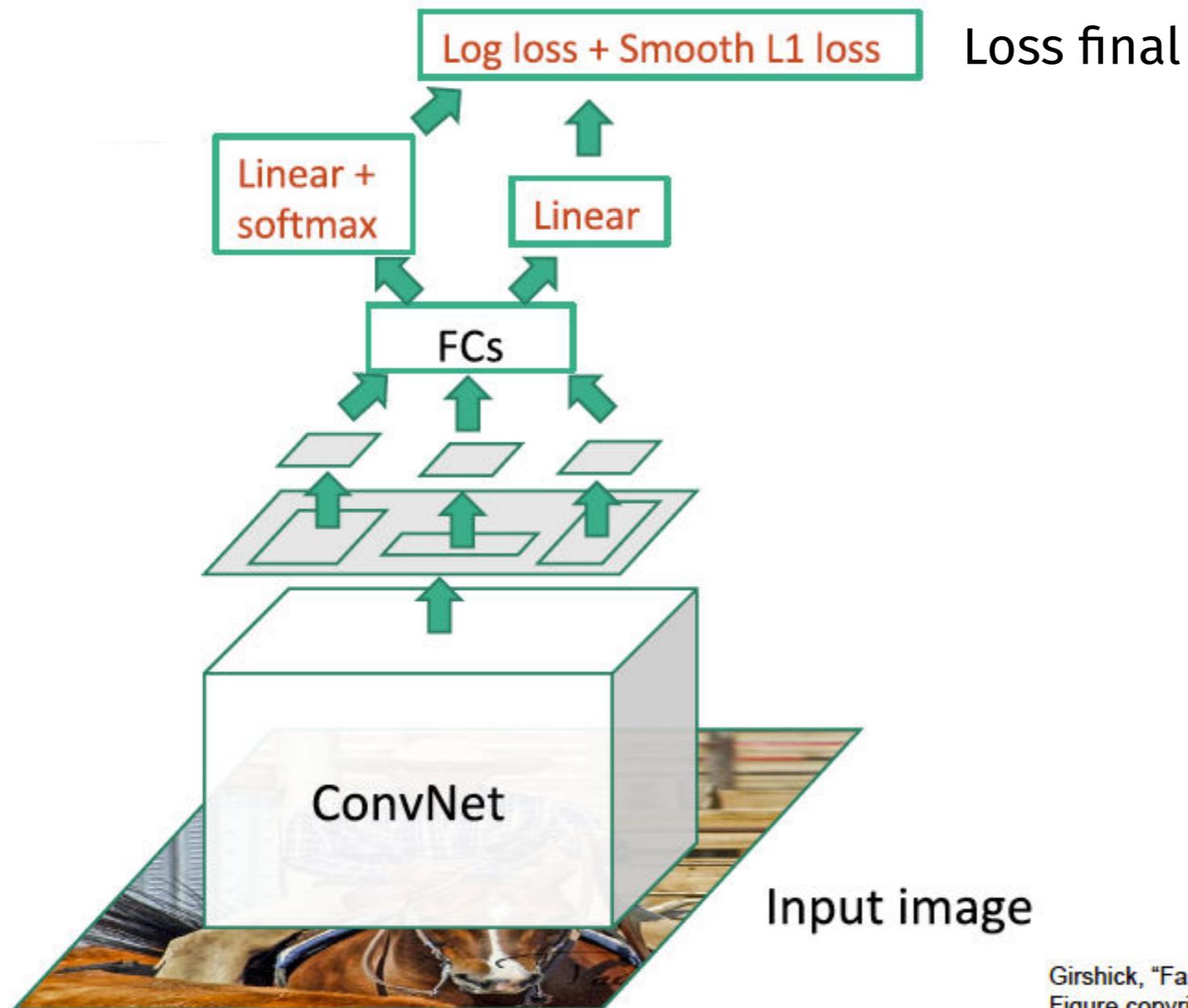


Girshick, “Fast R-CNN”, ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: Fast R-CNN

Treinamento

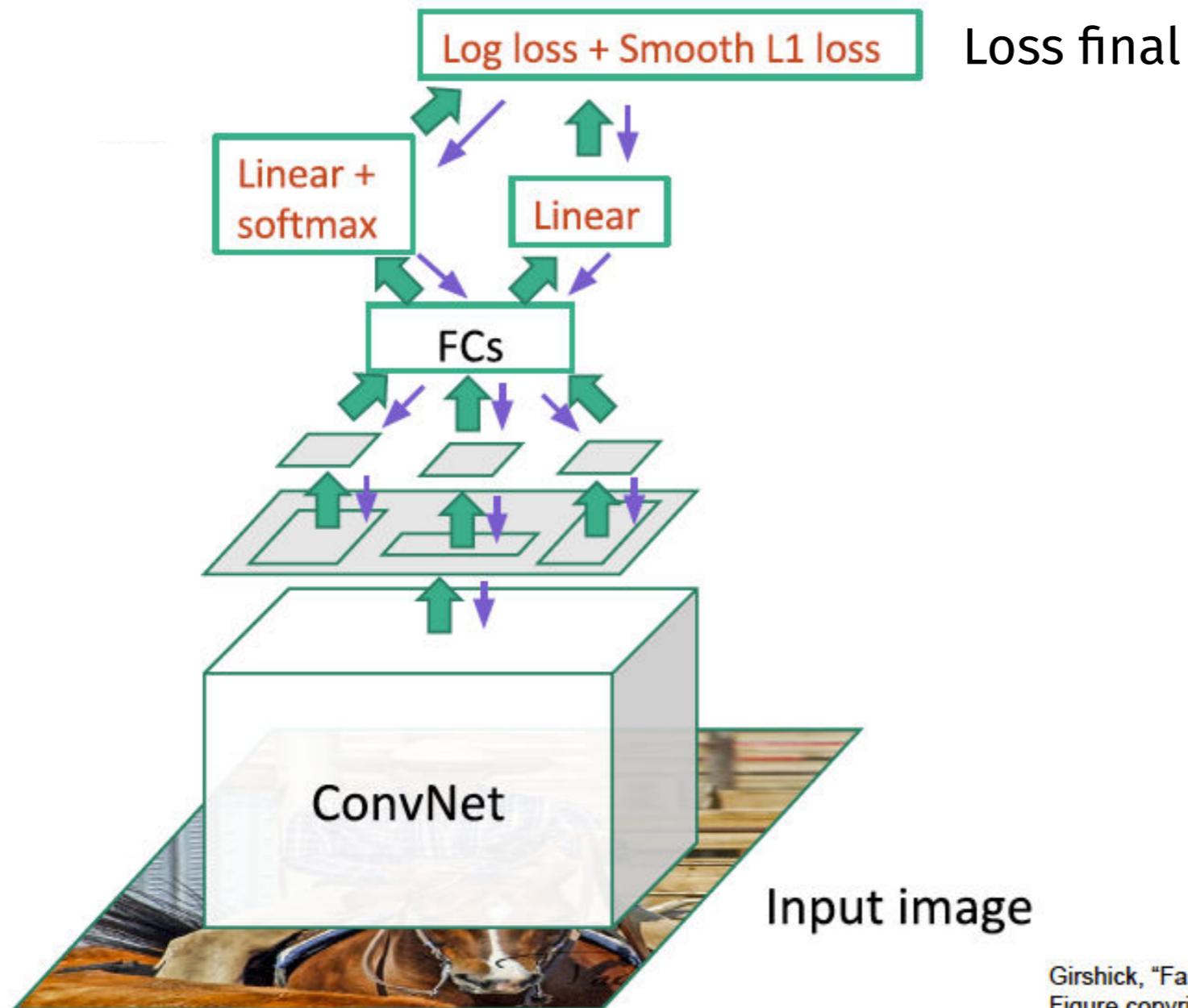


Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: Fast R-CNN

Treinamento

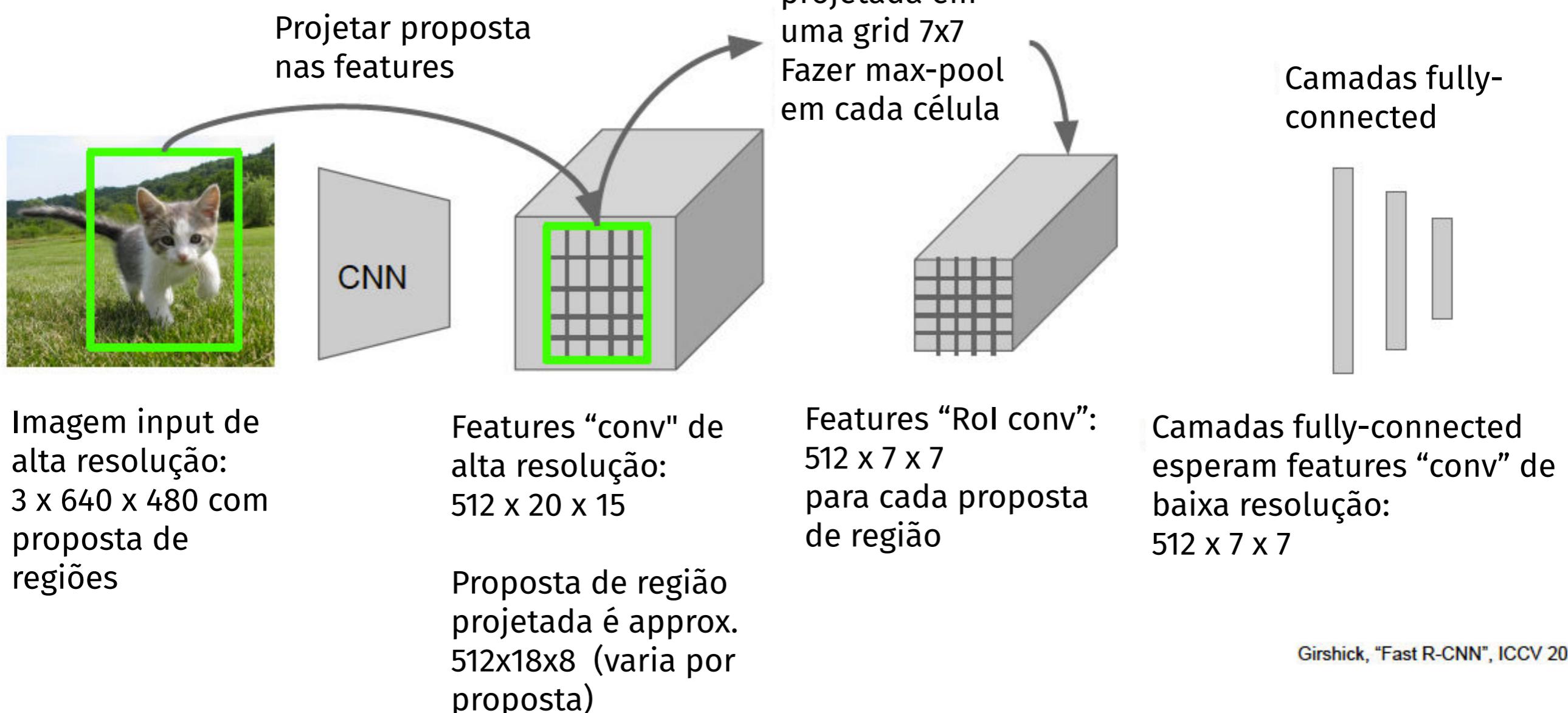


Girshick, "Fast R-CNN", ICCV 2015.
Figure copyright Ross Girshick, 2015; [source](#). Reproduced with permission.

Detecção e segmentação de objetos

Detecção de objetos: Fast R-CNN

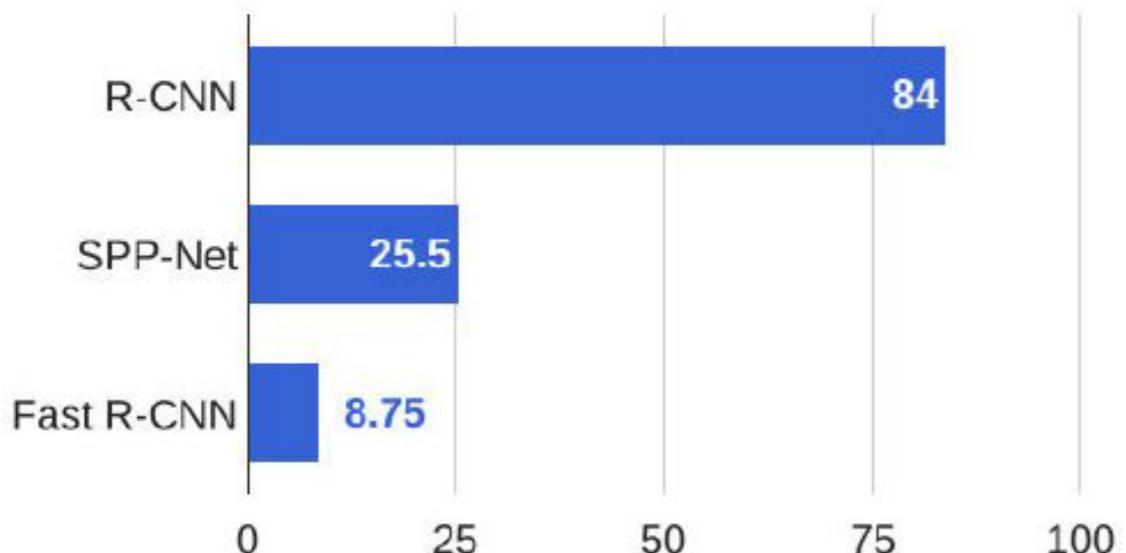
Rol pooling



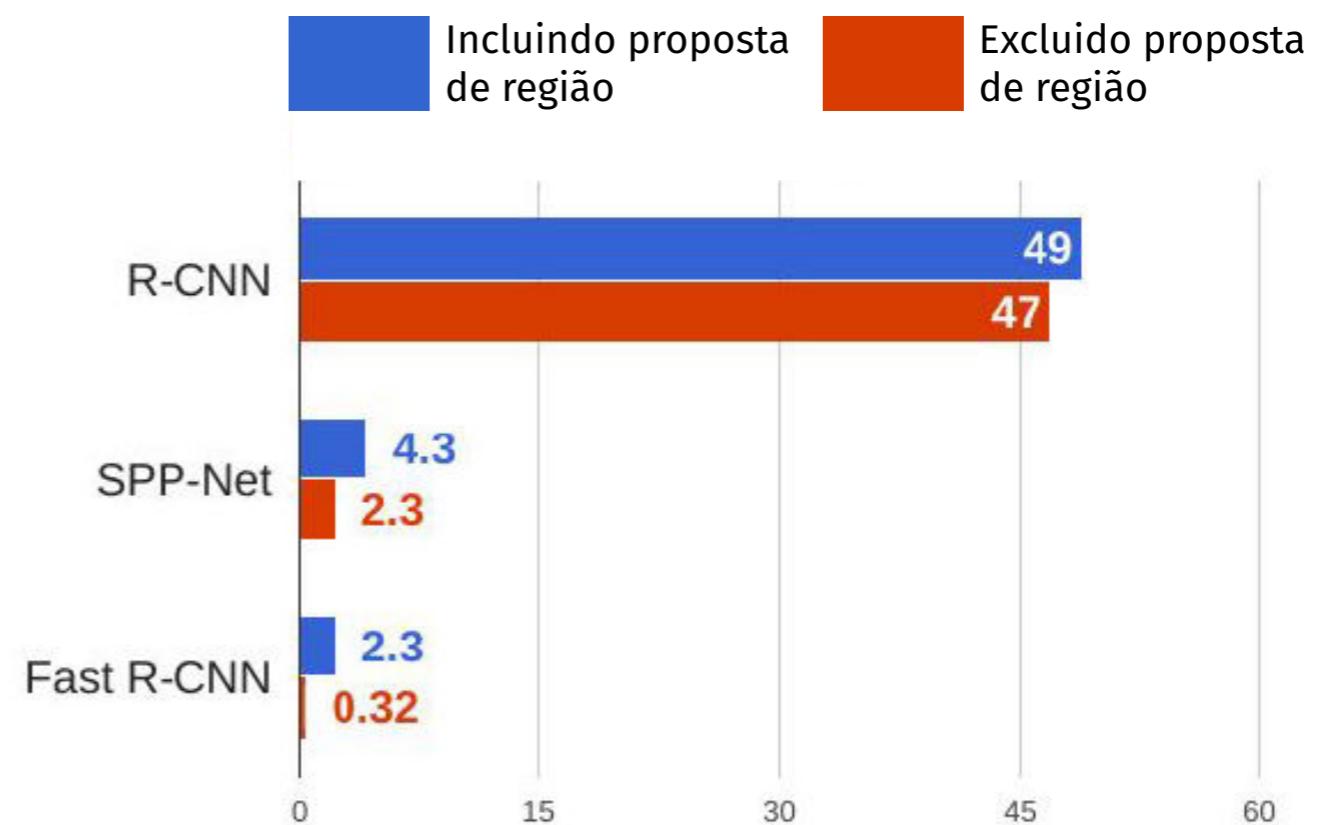
Detecção e segmentação de objetos

Detecção de objetos: R-CNN vs. SPP vs. Fast R-CNN

Tempo de treinamento (horas)



Tempo de teste (segundos)



Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.

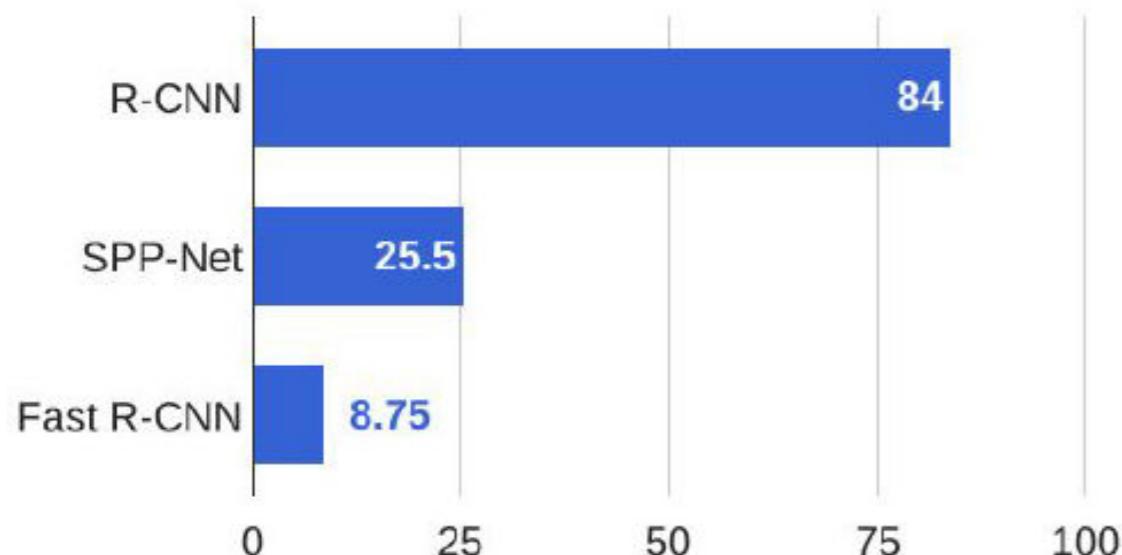
He et al, "Spatial pyramid pooling in deep convolutional networks for visual recognition", ECCV 2014

Girshick, "Fast R-CNN", ICCV 2015

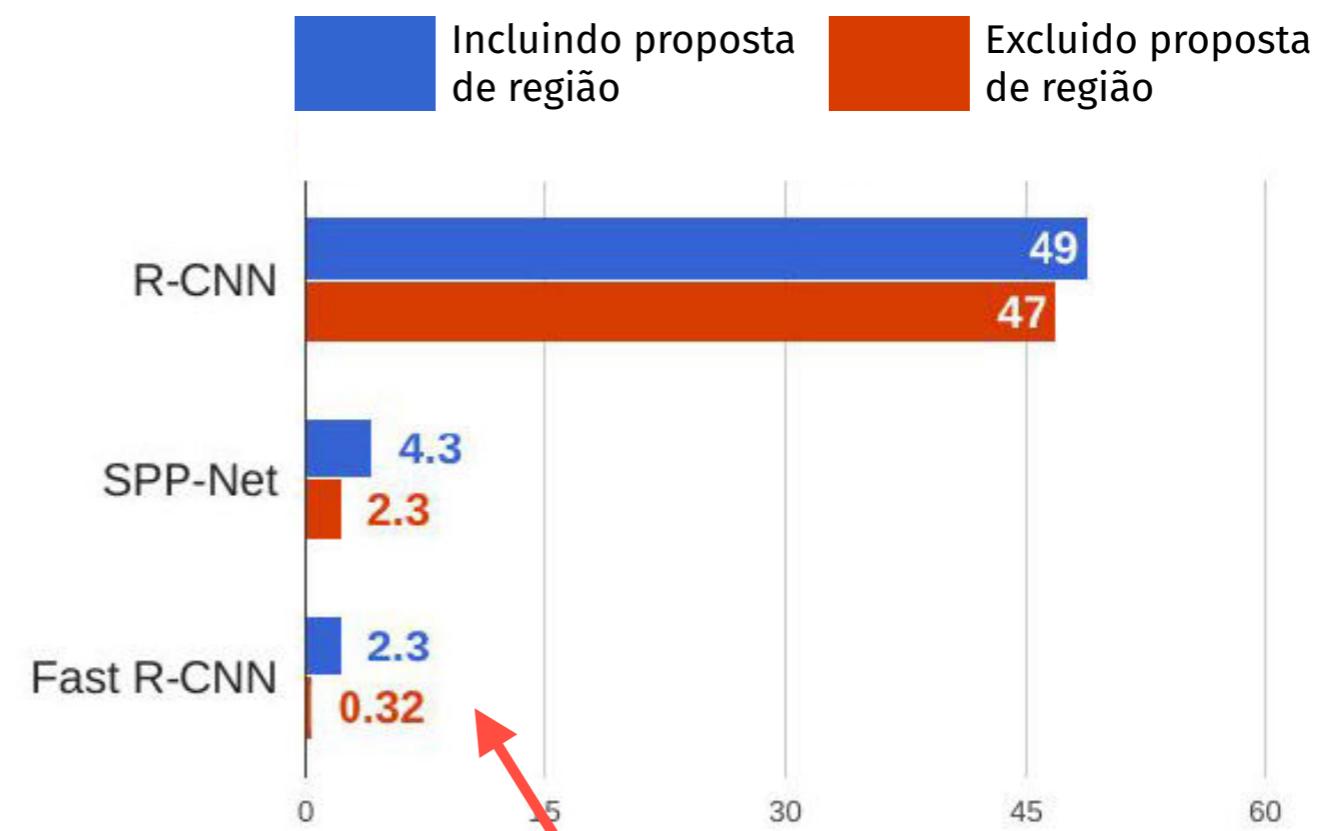
Detecção e segmentação de objetos

Detecção de objetos: R-CNN vs. SPP vs. Fast R-CNN

Tempo de treinamento (horas)



Tempo de teste (segundos)



Problema! Tempo dominado pela proposta de regiões.

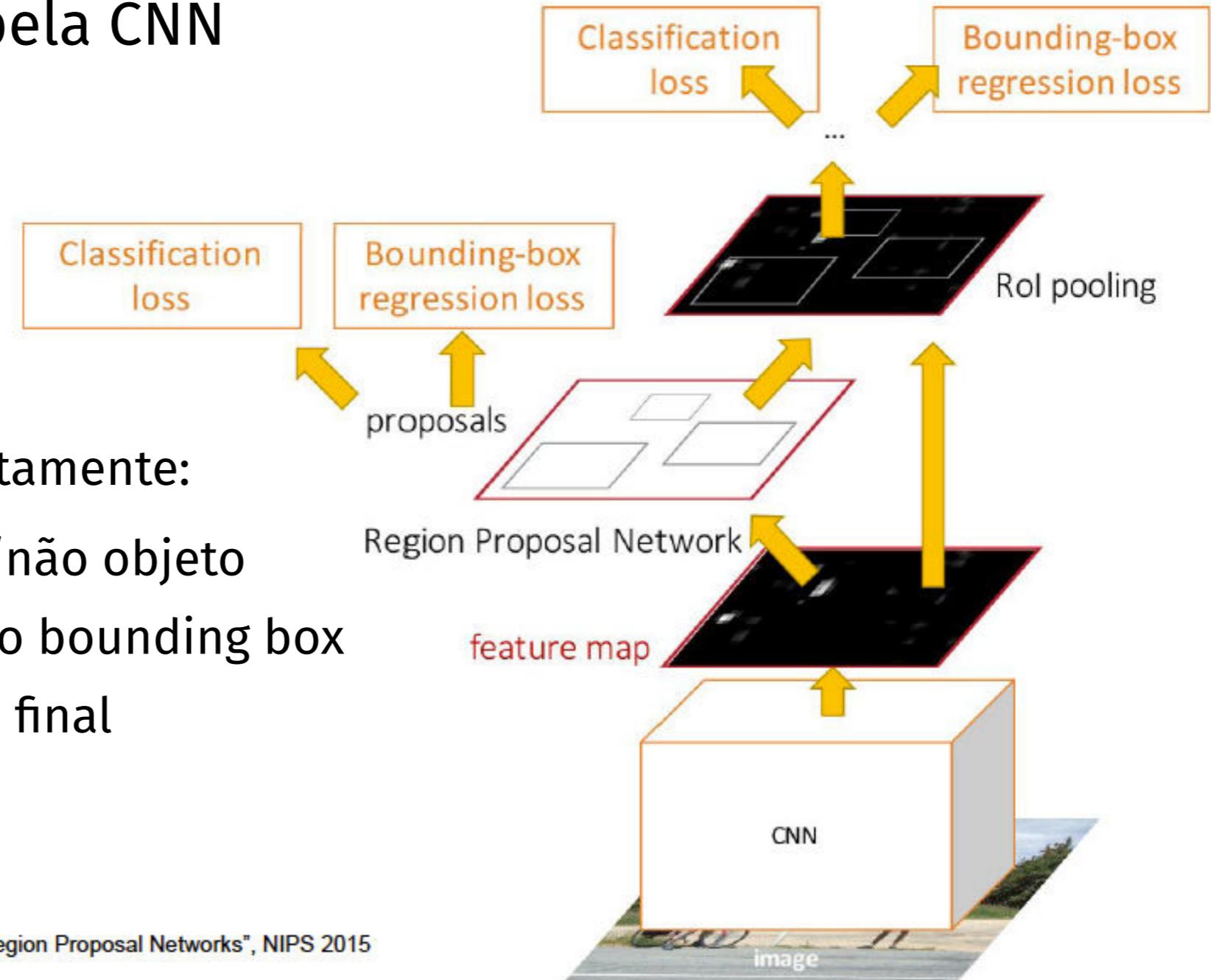
Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR 2014.
He et al, "Spatial pyramid pooling in deep convolutional networks for visual recognition", ECCV 2014
Girshick, "Fast R-CNN", ICCV 2015

Detecção e segmentação de objetos

Detecção de objetos: Faster R-CNN

Proposta de região pela CNN

- Inserir Region Proposal Network (RPN) para fazer propostas a partir das features
- Treinar 4 losses conjuntamente:
 - RPN classifica objeto/não objeto
 - RPN faz a regressão do bounding box
 - Score de classificação final
 - Bounding box final



Ren et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", NIPS 2015
Figure copyright 2015, Ross Girshick; reproduced with permission

Detecção e segmentação de objetos

Detecção de objetos: comparação



Detecção e segmentação de objetos

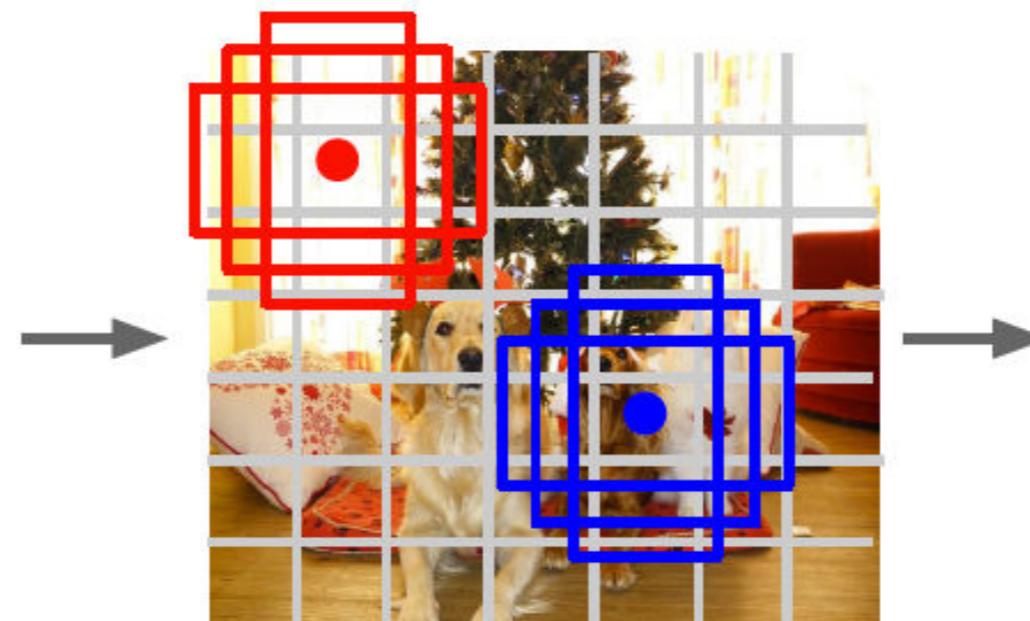
Detecção de objetos: YOLO / SSD

Sem proposta de região

Da imagem de entrada aos scores com uma rede convolucional grande



Imagen de entrada
 $3 \times H \times W$



Dividir a imagem em
um grid 7×7

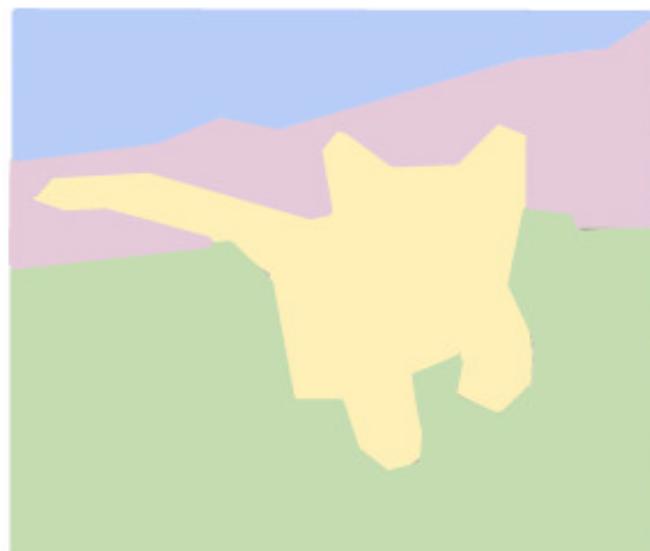
Criar um conjunto de boxes
base centrados em cada célula

- Para cada célula do grid:
- Fazer regressão a partir das **B** boxes base para a box final: (dx , dy , dh , dw , confiança)
 - Predizer scores para cada uma das **C** classes (incluindo o background)
 - Output: $7 \times 7 \times (5 * B + C)$

Redmon et al, "You Only Look Once:
Unified, Real-Time Object Detection", CVPR 2016
Liu et al, "SSD: Single-Shot MultiBox Detector", ECCV 2016

Detecção e segmentação de objetos

Segmentação
semântica
(apenas pixels)



**GRASS, CAT,
TREE, SKY**

Classificação +
Localização
(objeto único)



CAT

Detecção
(objetos múltiplos)



DOG, DOG, CAT

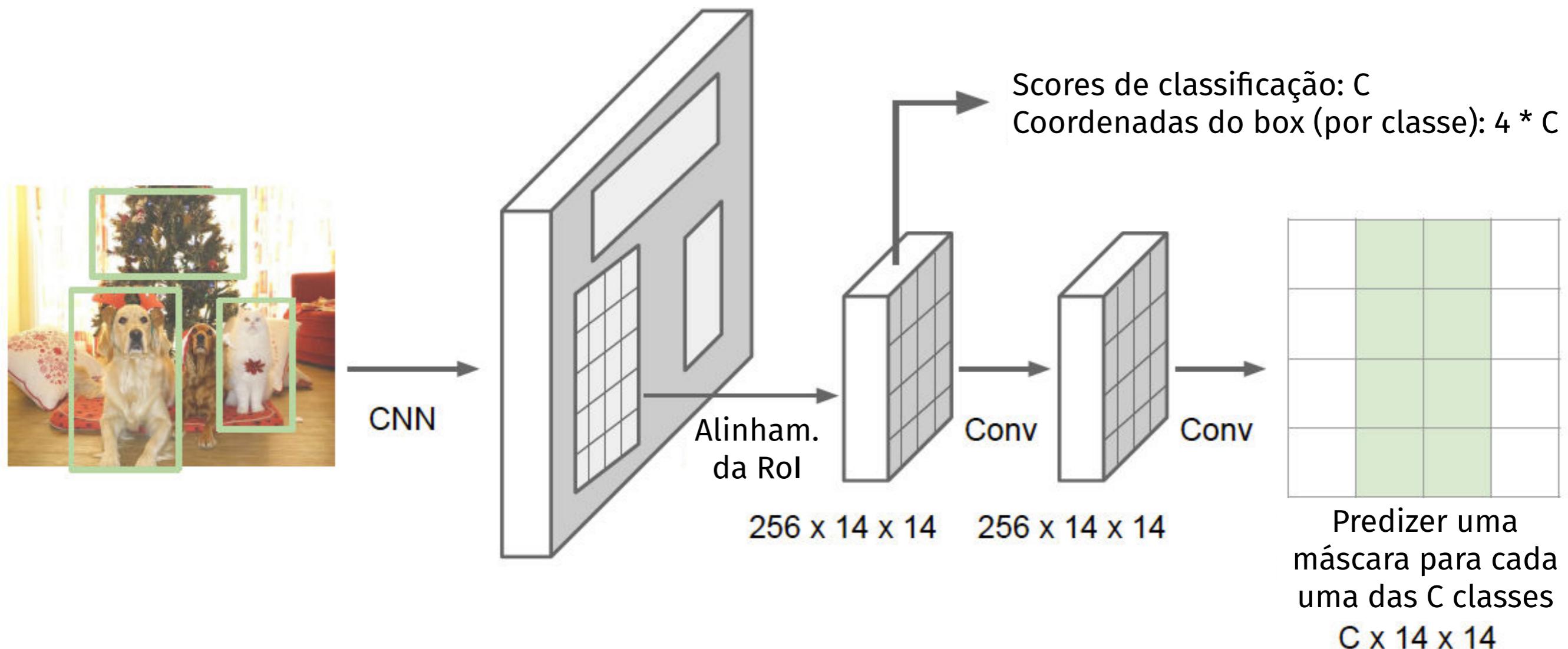
Segmentação de
instâncias
(objetos múltiplos)



DOG, DOG, CAT

Detecção e segmentação de objetos

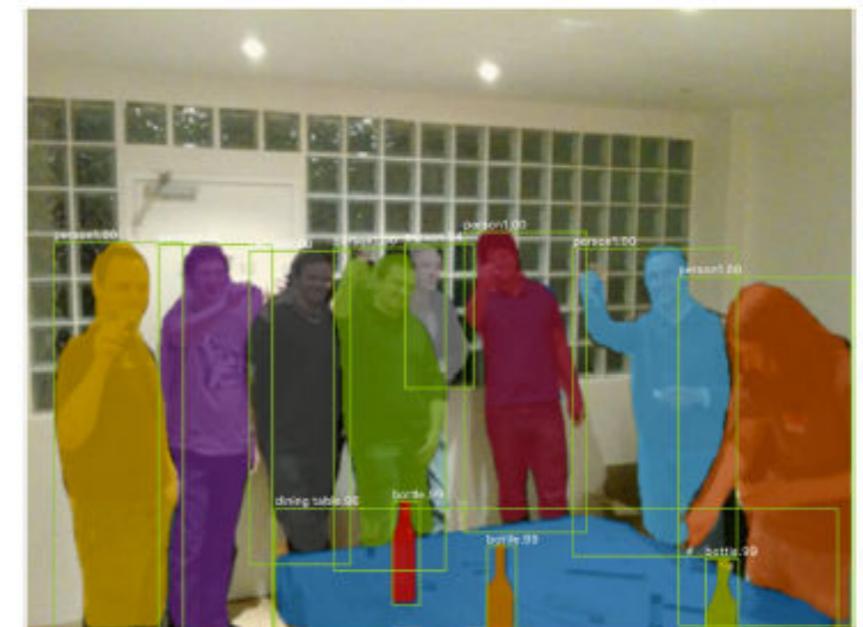
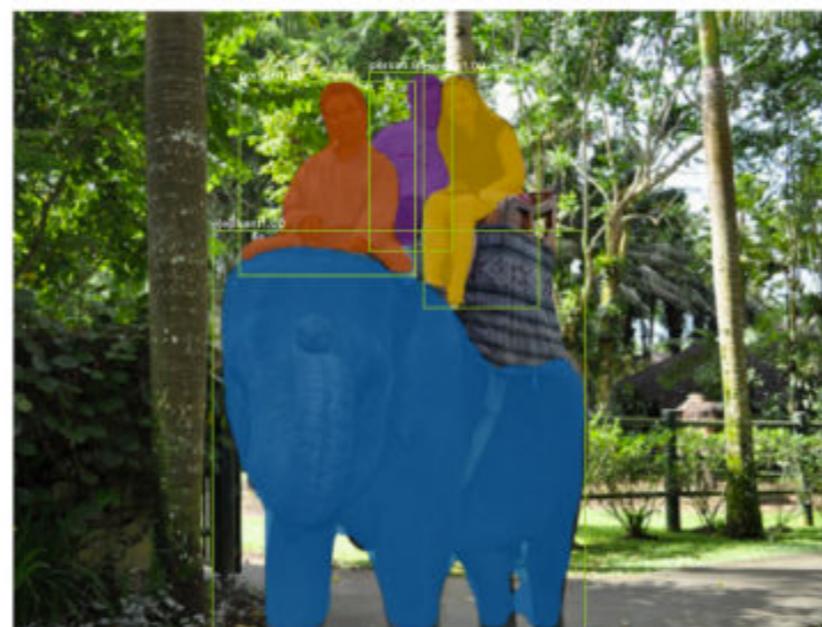
Segmentação de instâncias: Mask R-CNN



He et al, "Mask R-CNN", arXiv 2017

Detecção e segmentação de objetos

Segmentação de instâncias: Mask R-CNN



He et al, "Mask R-CNN", arXiv 2017

Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.

Reproduced with permission.

Detecção e segmentação de objetos

Segmentação de instâncias: Mask R-CNN

Exemplo: Estimativa de pose em humanos



He et al, "Mask R-CNN", arXiv 2017
Figures copyright Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 2017.
Reproduced with permission.

Detecção e segmentação de objetos

Segmentação
semântica
(apenas pixels)



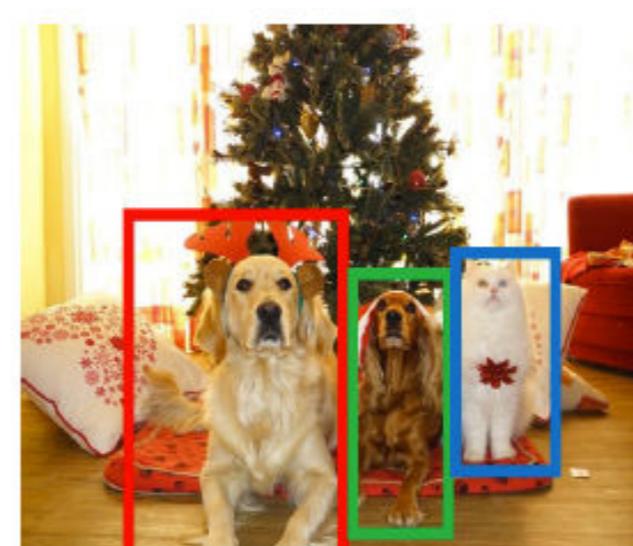
**GRASS, CAT,
TREE, SKY**

Classificação +
Localização
(objeto único)



CAT

Detecção
(objetos múltiplos)



DOG, DOG, CAT

Segmentação de
instâncias
(objetos múltiplos)



DOG, DOG, CAT

Transferência de estilo

Transferência de estilo

O que é transferência de estilo?



Conteúdo (C)

The Great Wave off Kanagawa
Hokusai



Estilo (S)

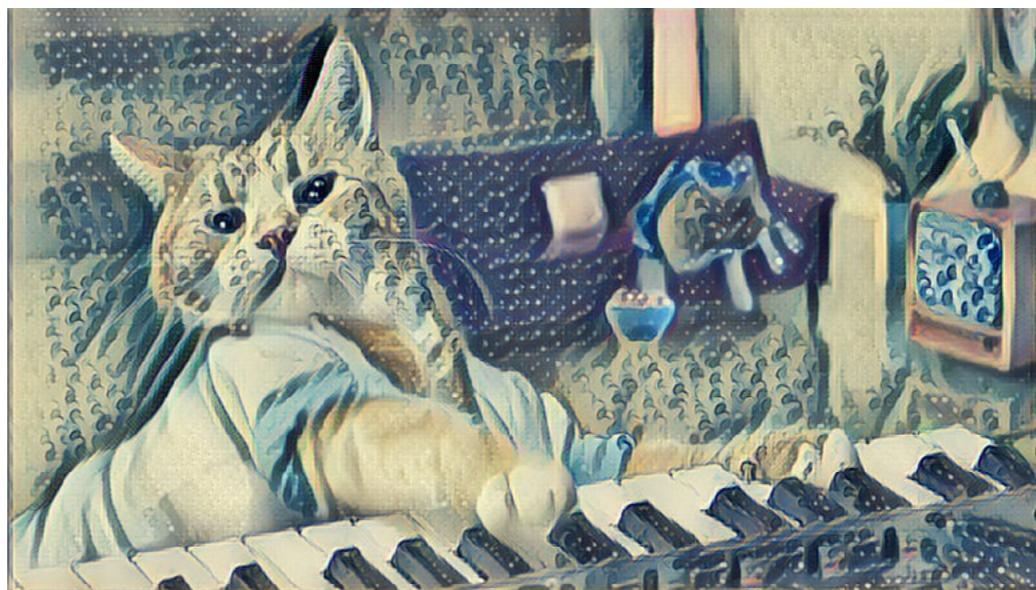


Imagen gerada (G)

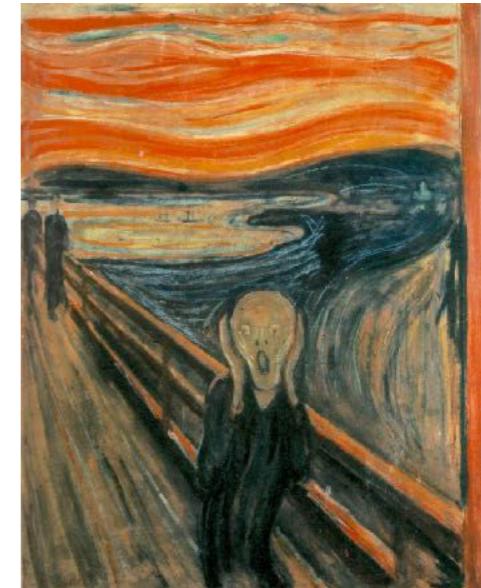
Transferência de estilo

O que é transferência de estilo?



Conteúdo (C)

The Scream
Edvard Munch



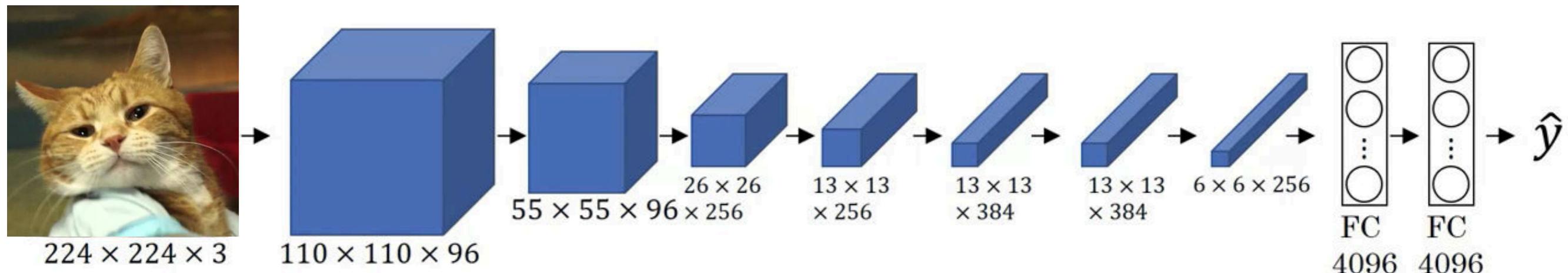
Estilo (S)



Imagen gerada (G)

Transferência de estilo

Intuição: o que CNNs aprendem?

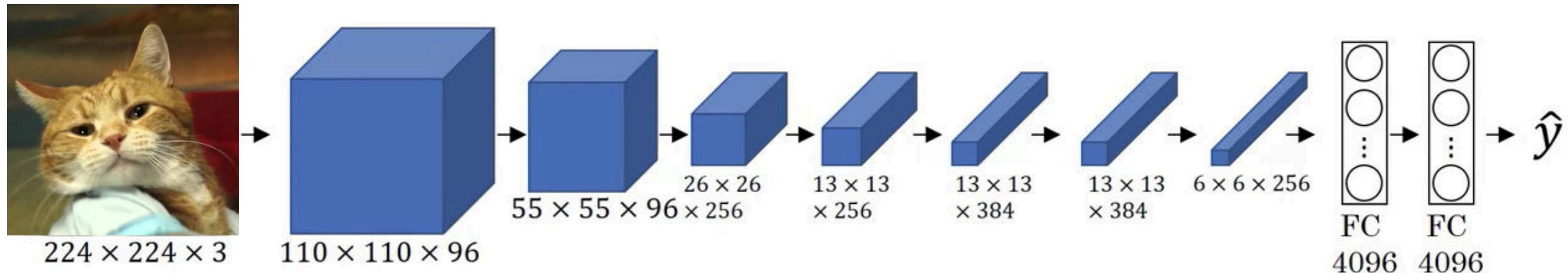


- Selecionar uma unidade na camada 1.

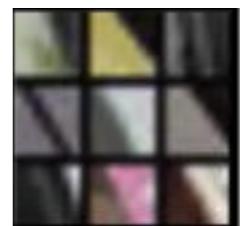
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



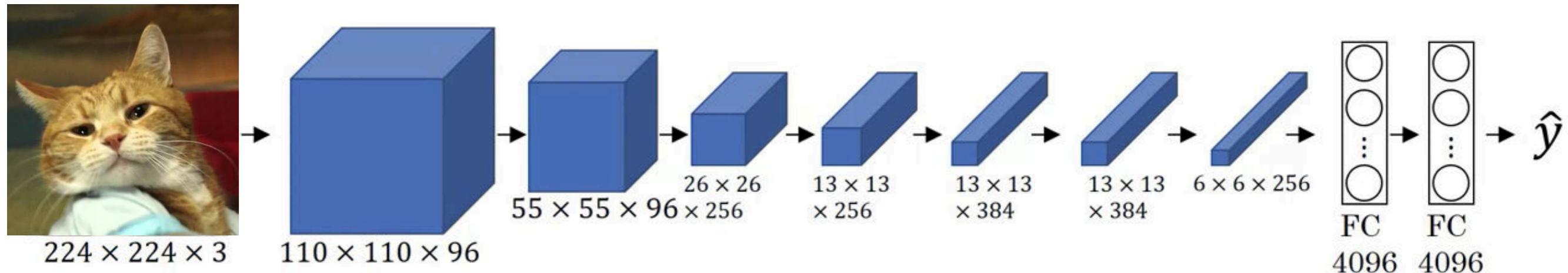
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.



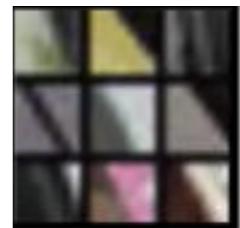
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



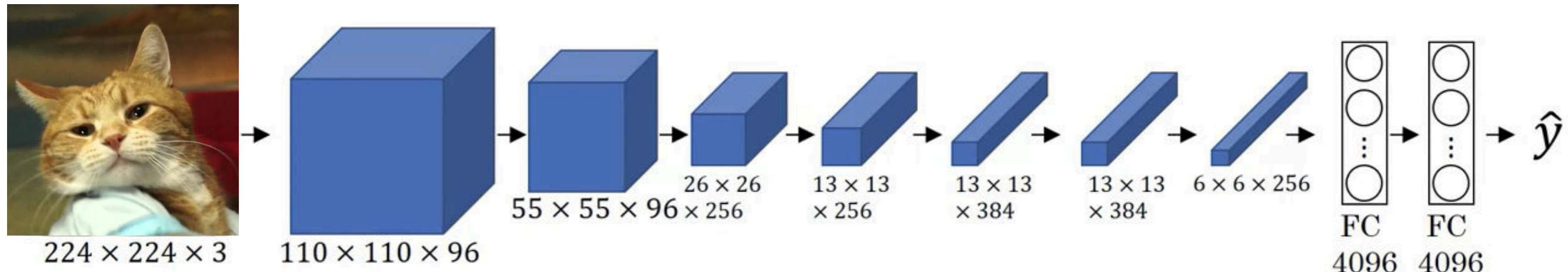
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



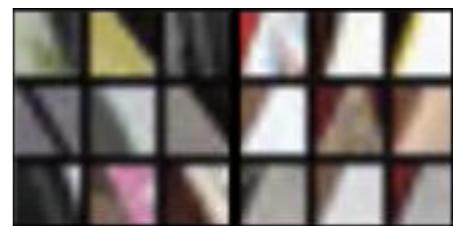
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



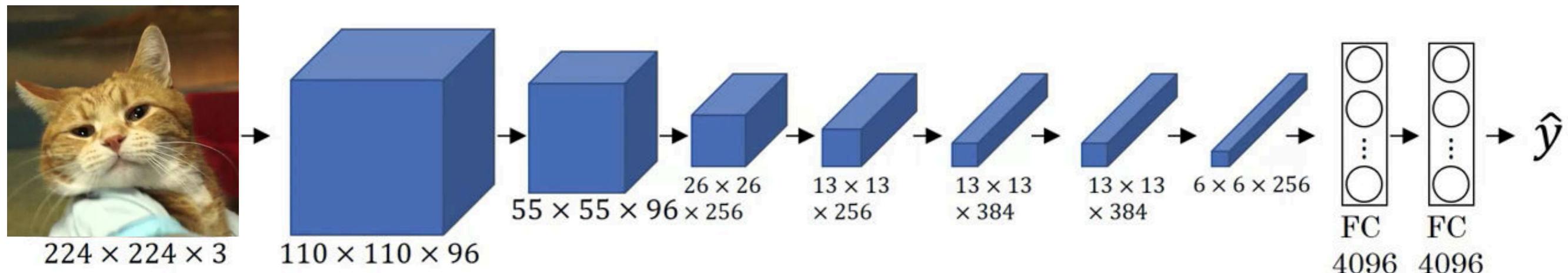
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



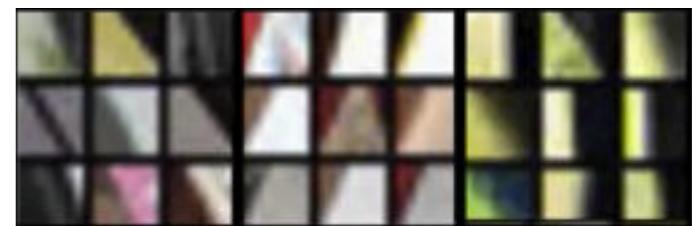
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



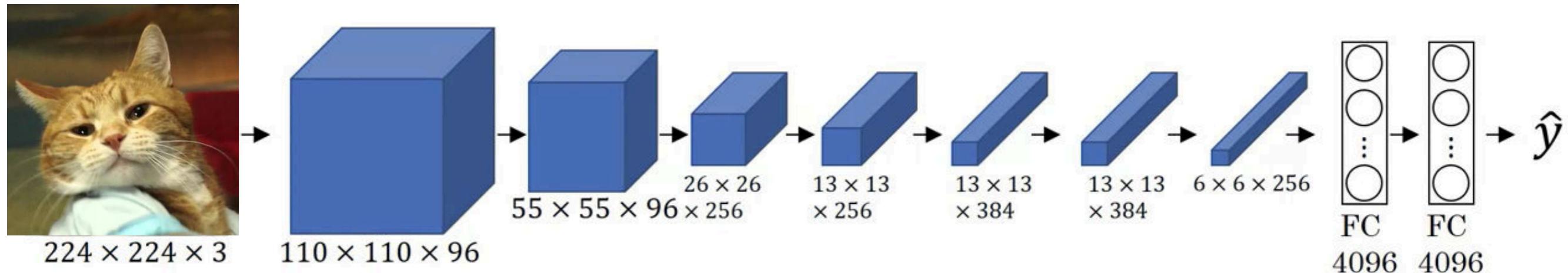
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



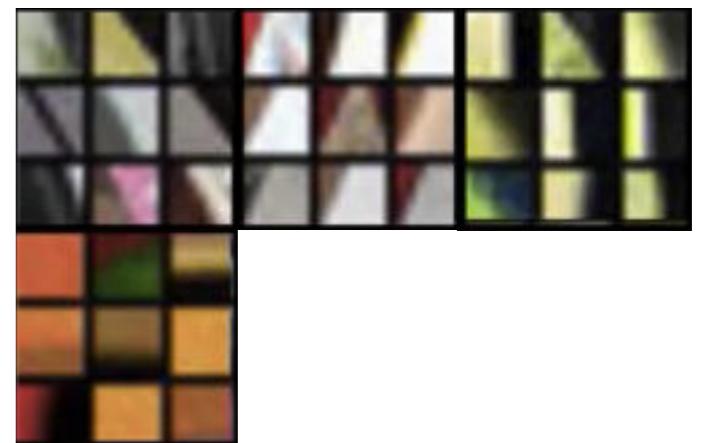
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



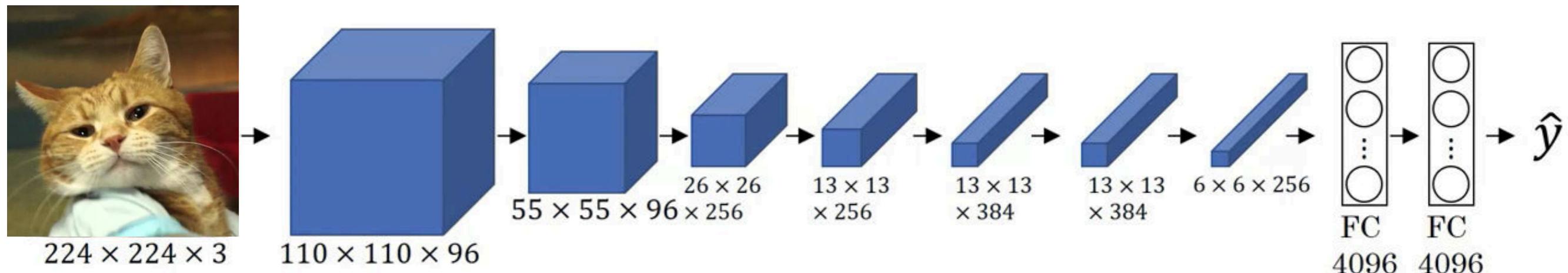
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



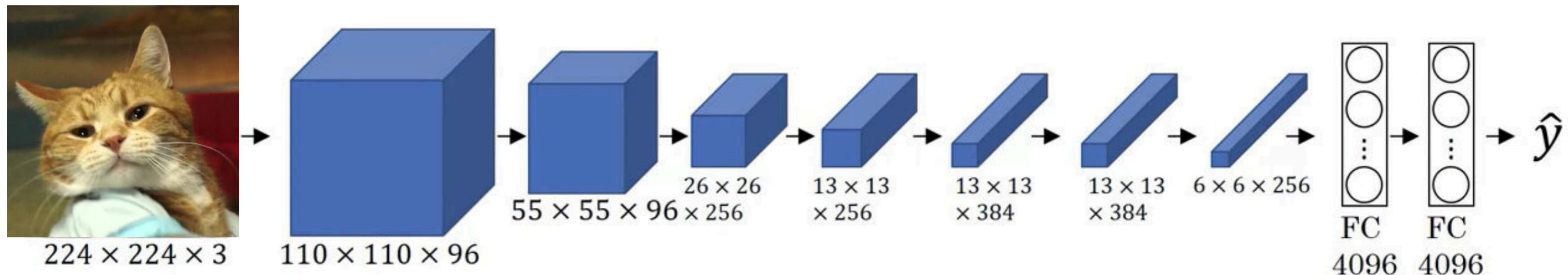
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



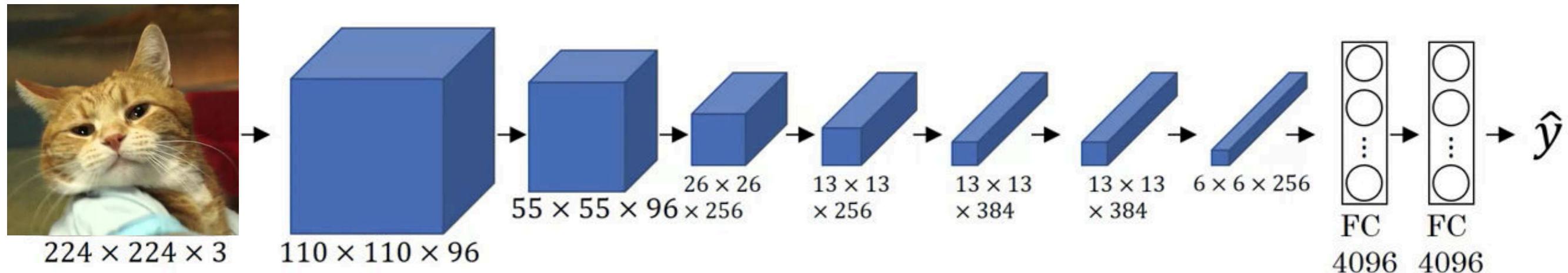
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



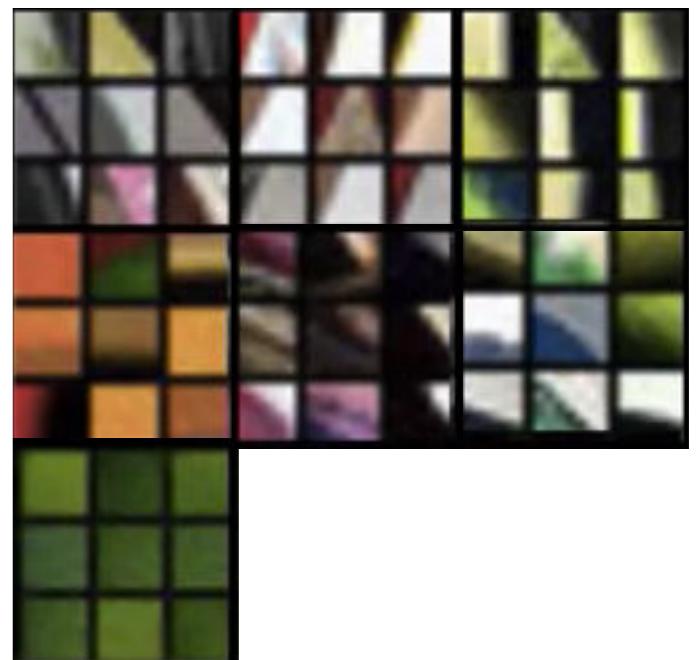
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



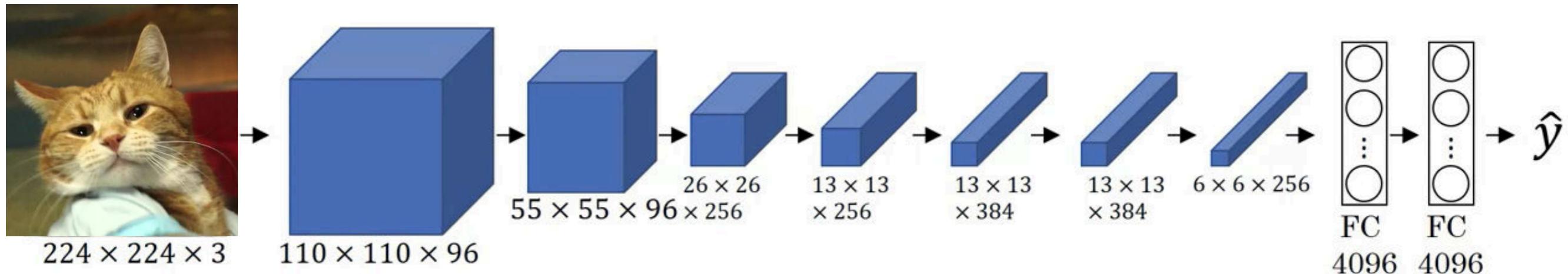
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



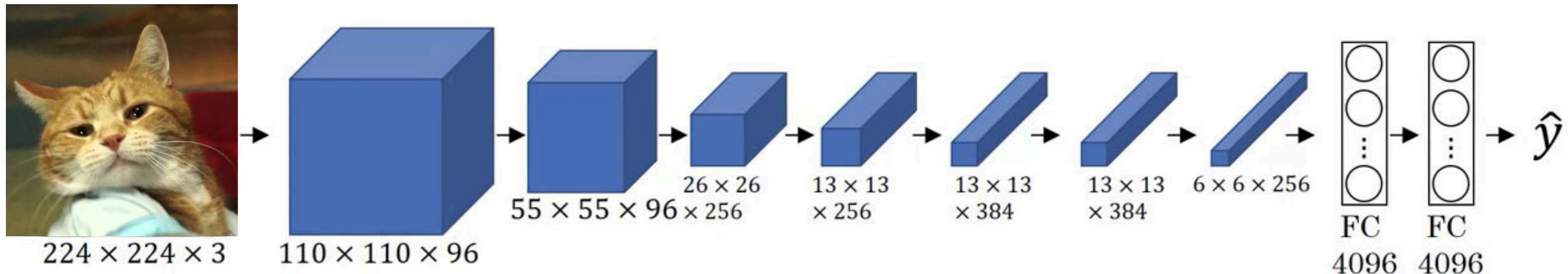
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



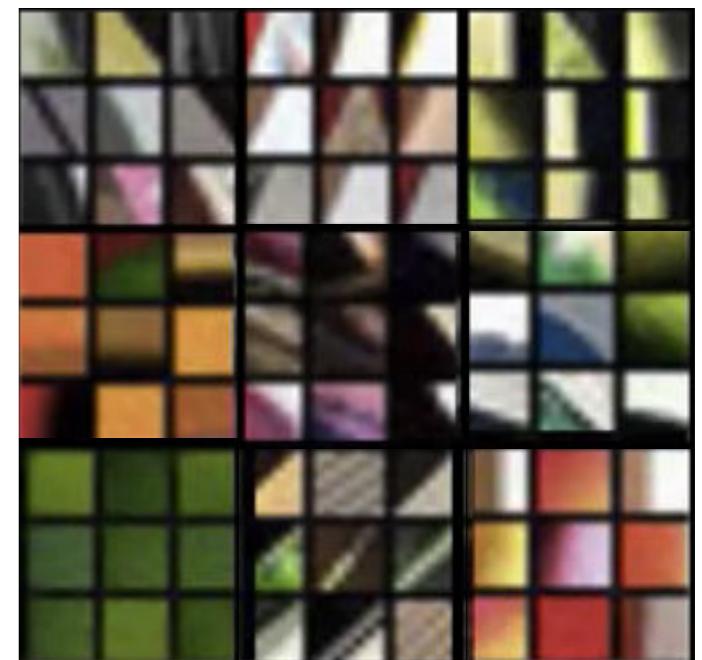
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



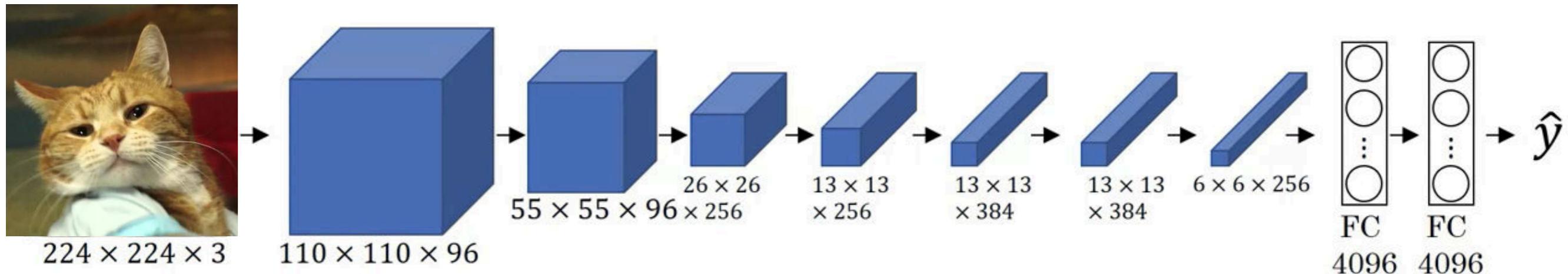
- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.



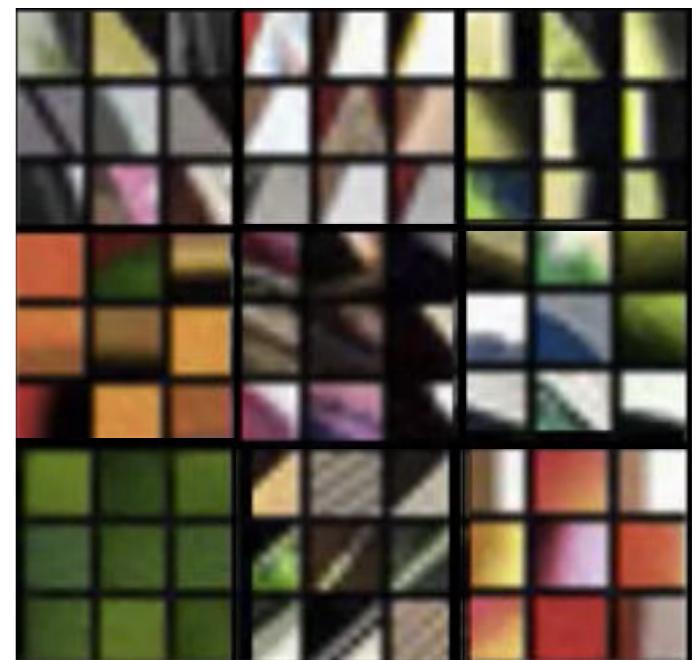
[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?



- Selecionar uma unidade na camada 1.
- Procurar os 9 patches de imagem que maximizam a ativação da unidade.
- Repetir para outras unidades.
- Repetir para camadas mais profundas.

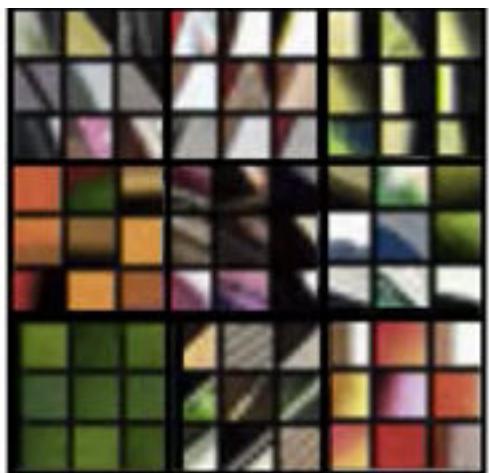


[Zeiler and Fergus, 2013, Visualizing and understanding convolutional networks]

Transferência de estilo

Intuição: o que CNNs aprendem?

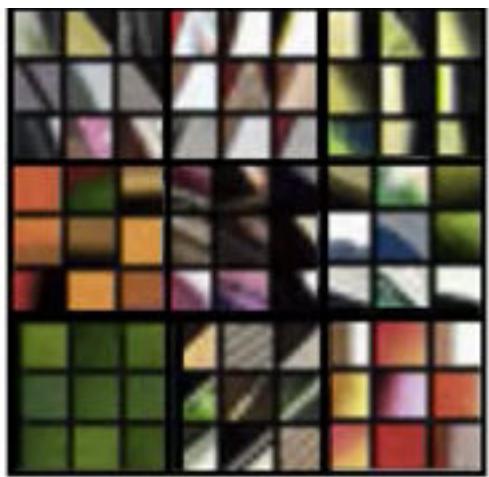
Layer 1



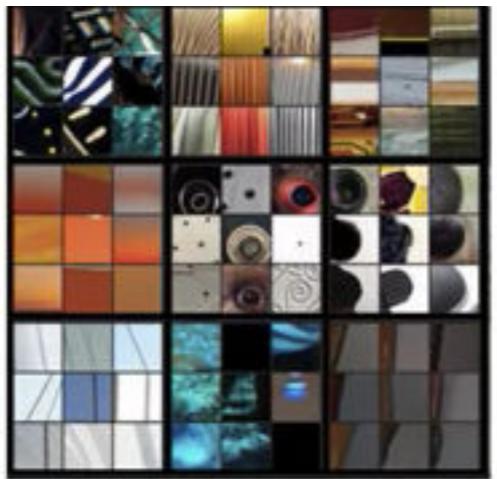
Transferência de estilo

Intuição: o que CNNs aprendem?

Layer 1



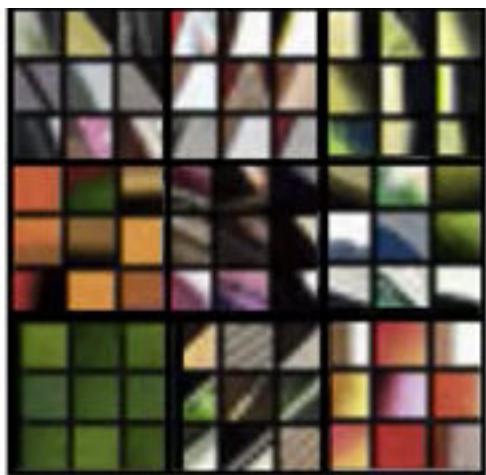
Layer 2



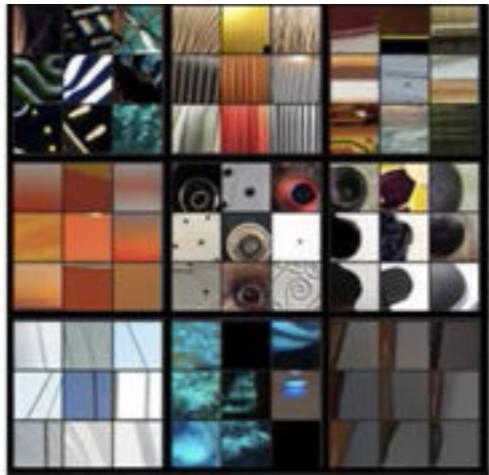
Transferência de estilo

Intuição: o que CNNs aprendem?

Layer 1



Layer 2



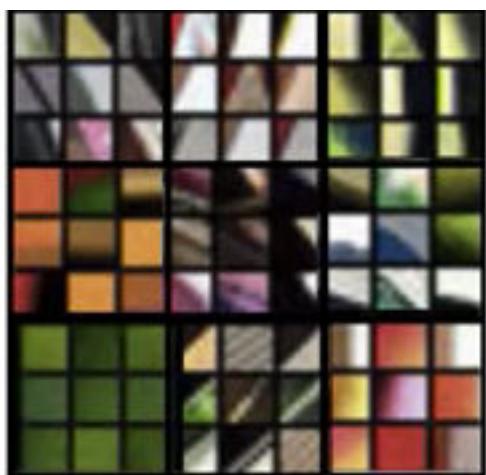
Layer 3



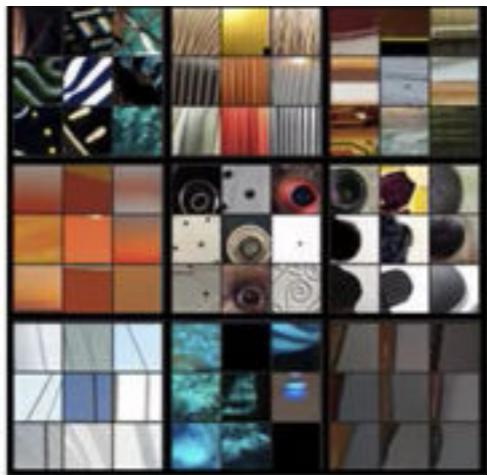
Transferência de estilo

Intuição: o que CNNs aprendem?

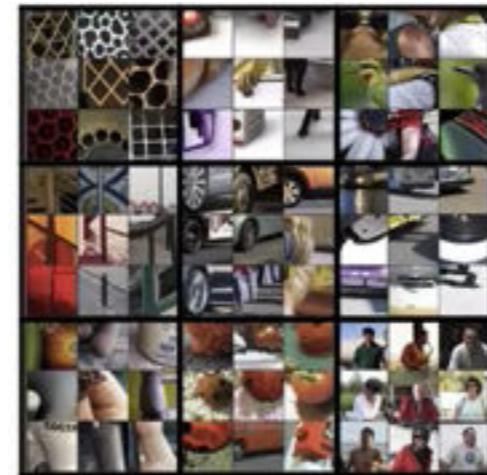
Layer 1



Layer 2



Layer 3



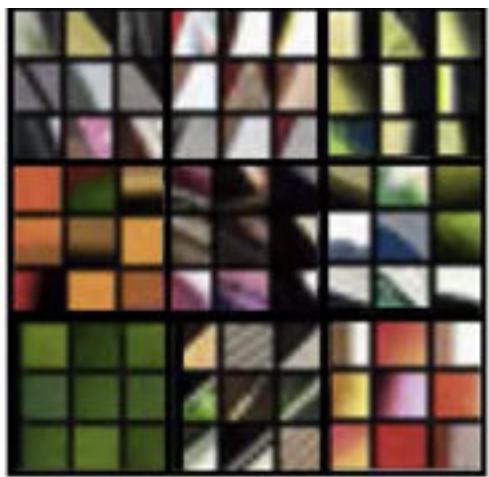
Layer 4



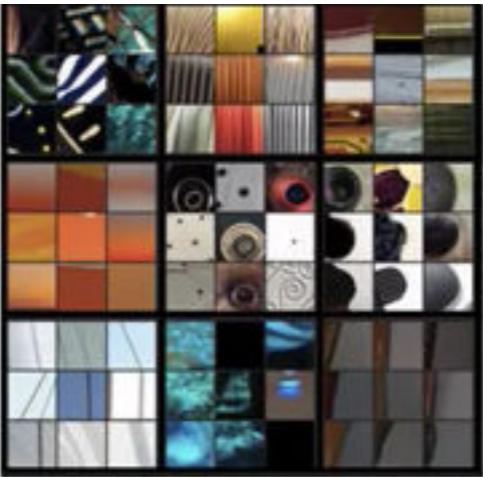
Transferência de estilo

Intuição: o que CNNs aprendem?

Layer 1



Layer 2



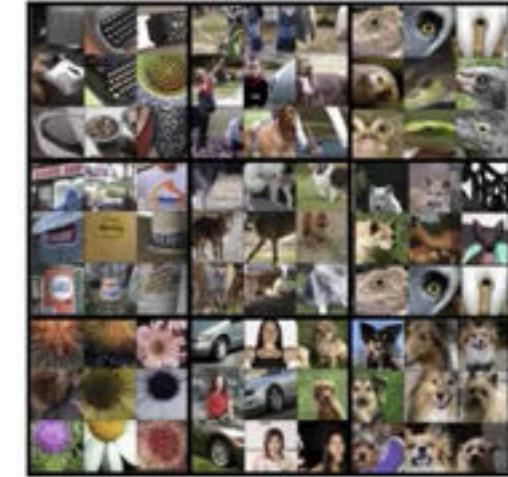
Layer 3



Layer 4



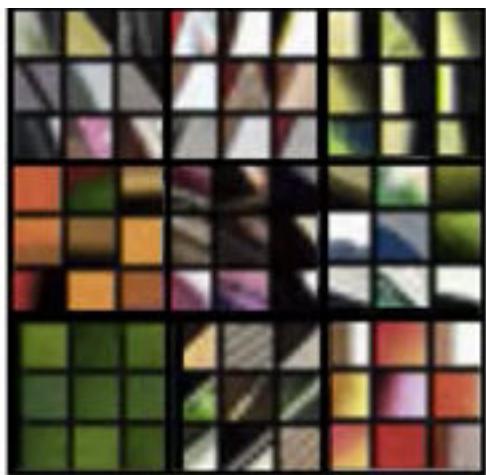
Layer 5



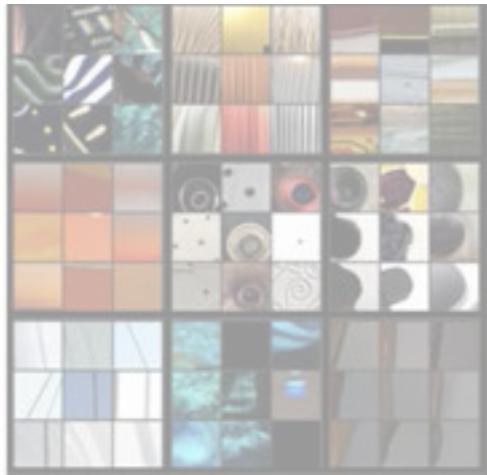
Transferência de estilo

Intuição: o que CNNs aprendem?

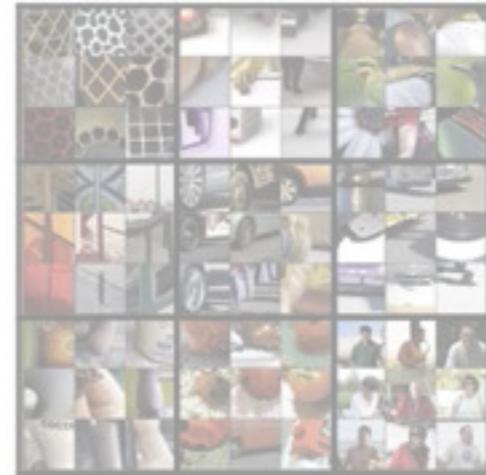
Layer 1



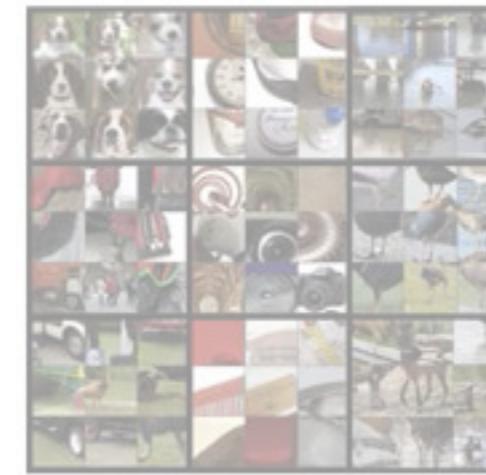
Layer 2



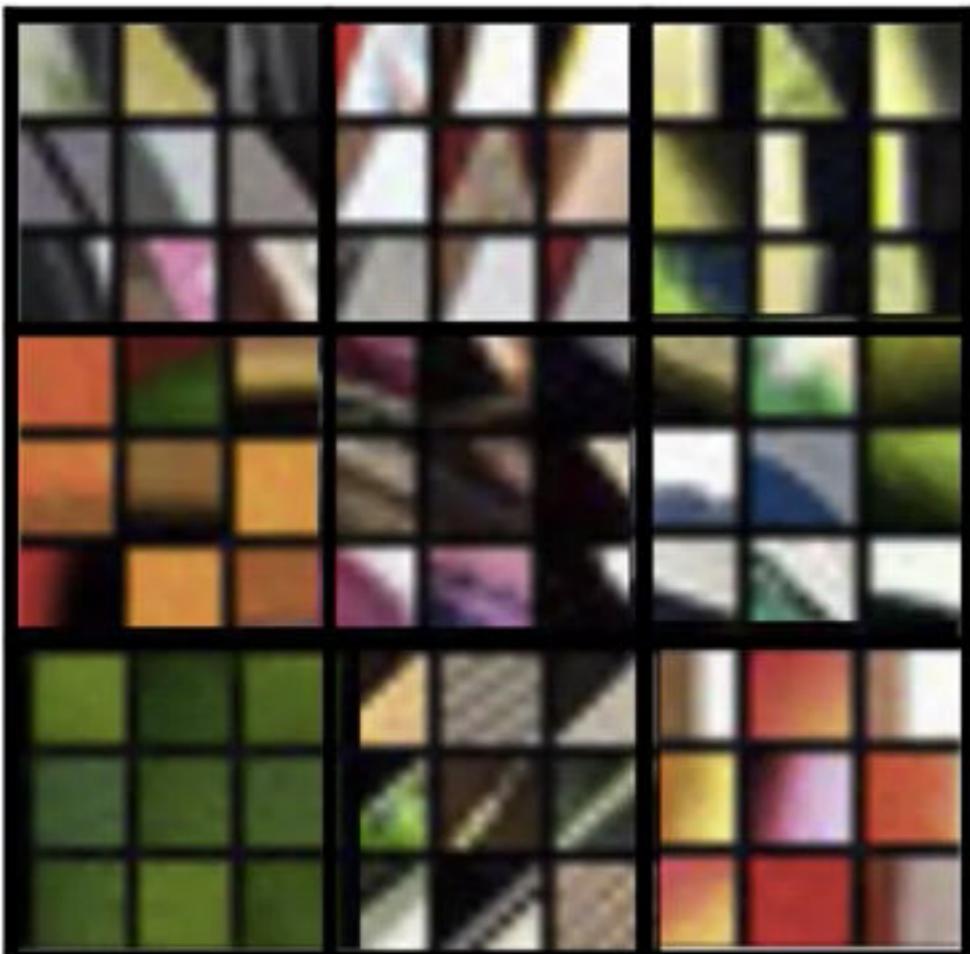
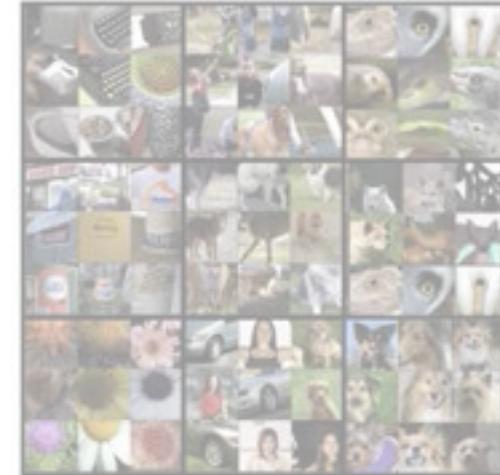
Layer 3



Layer 4



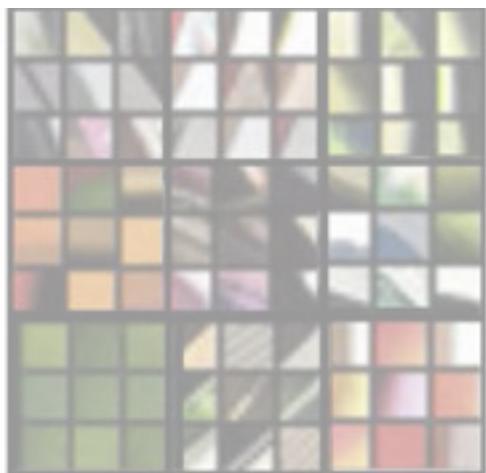
Layer 5



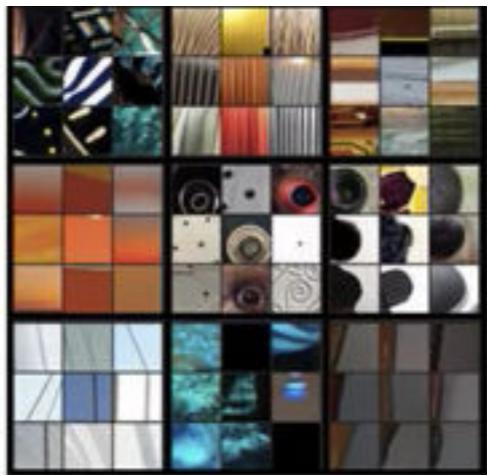
Transferência de estilo

Intuição: o que CNNs aprendem?

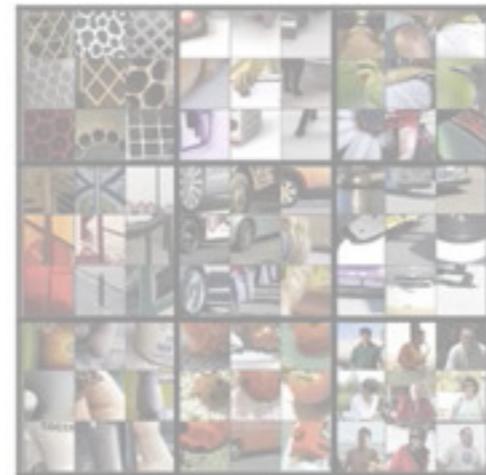
Layer 1



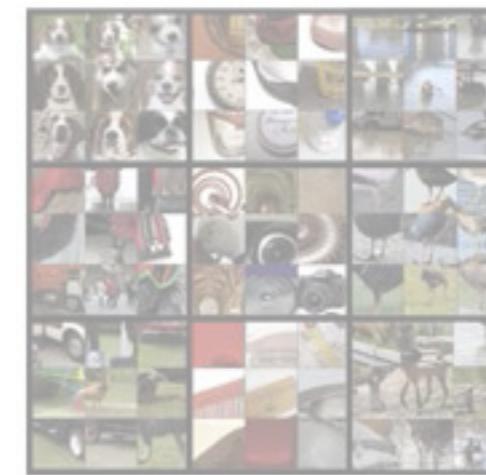
Layer 2



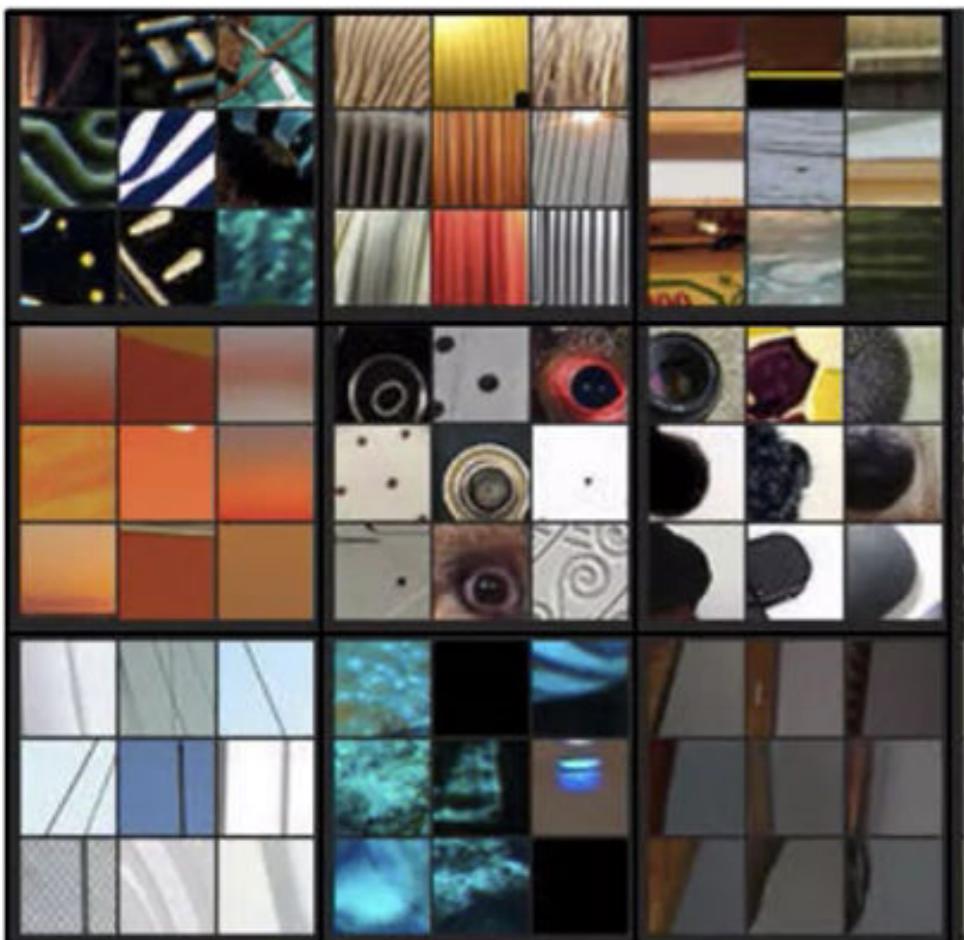
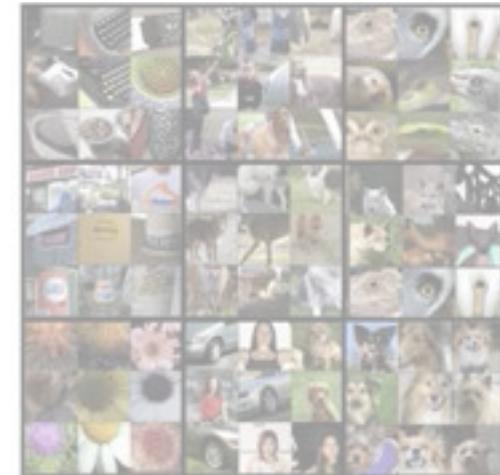
Layer 3



Layer 4



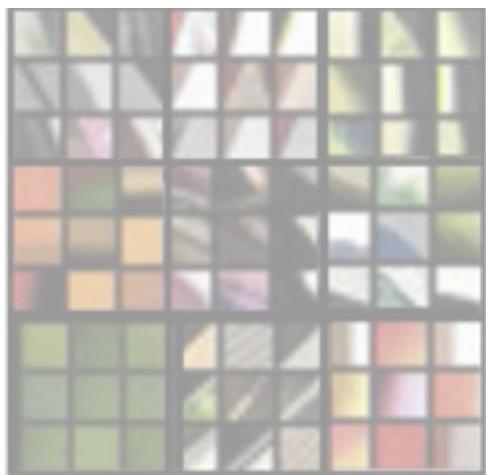
Layer 5



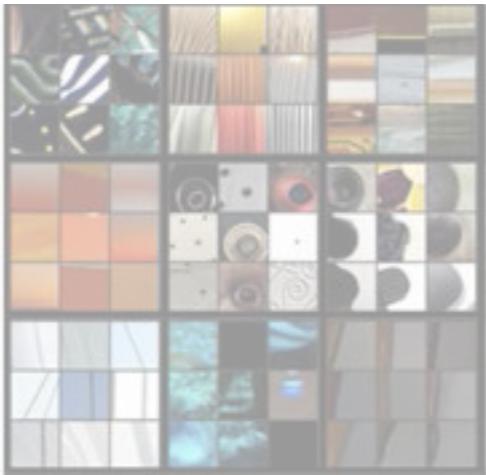
Transferência de estilo

Intuição: o que CNNs aprendem?

Layer 1



Layer 2



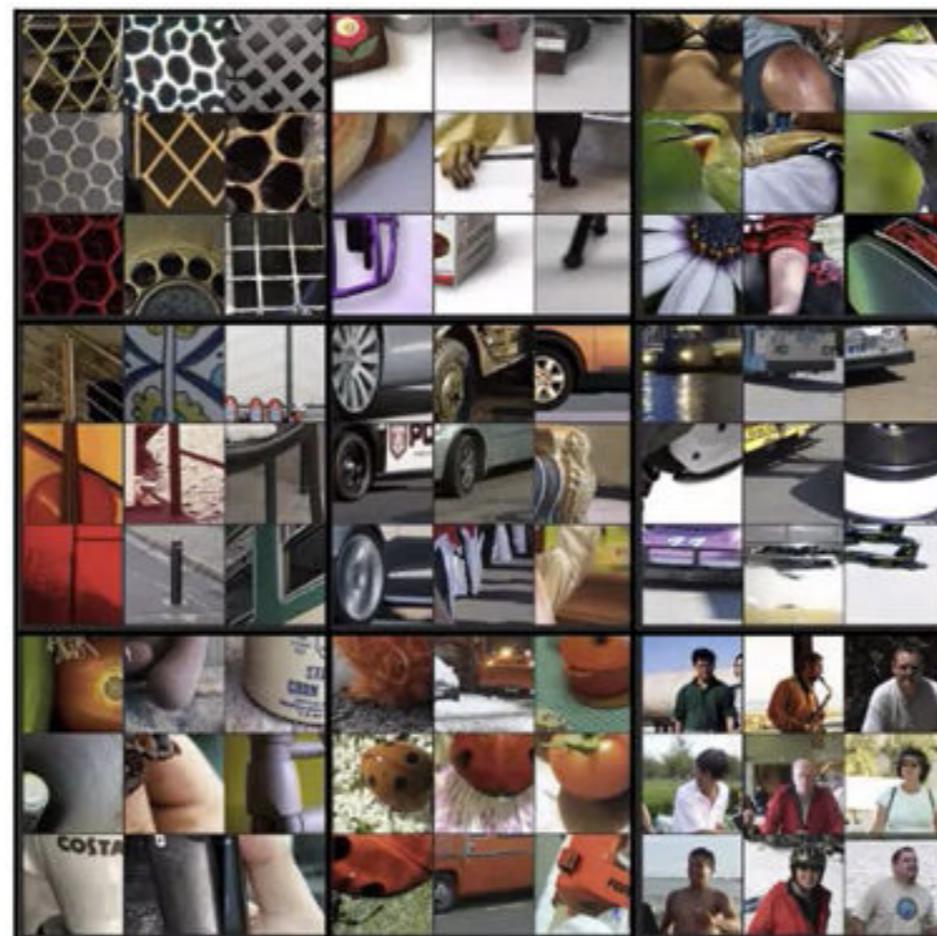
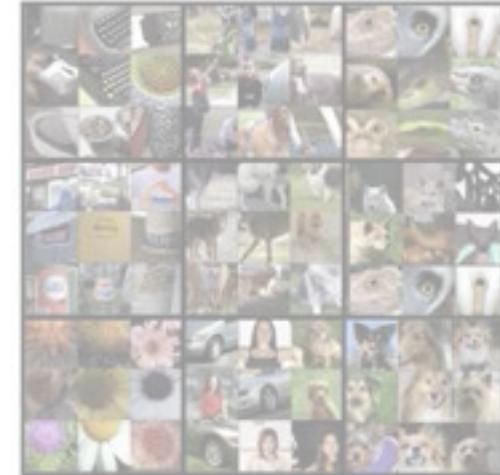
Layer 3



Layer 4



Layer 5



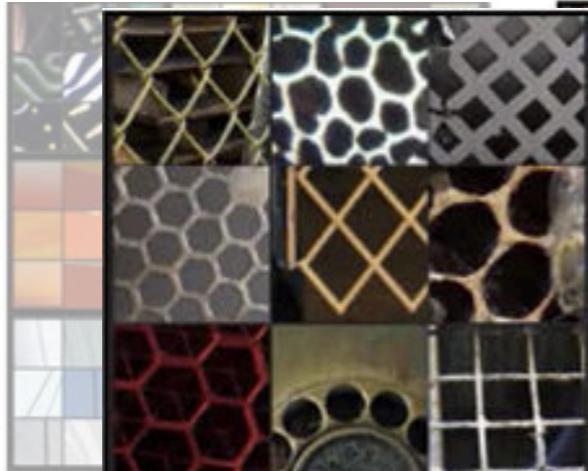
Transferência de estilo

Intuição: o que CNNs aprendem?

Layer 1



Layer 2



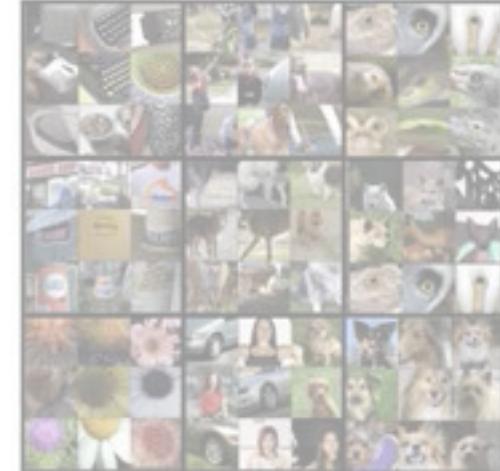
Layer 3



Layer 4



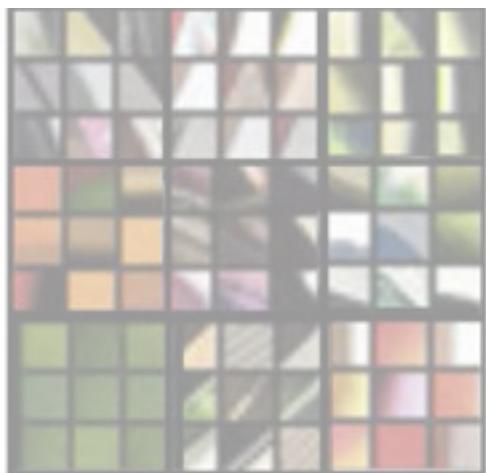
Layer 5



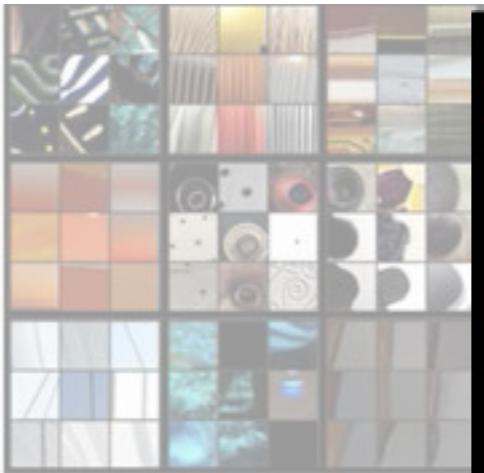
Transferência de estilo

Intuição: o que CNNs aprendem?

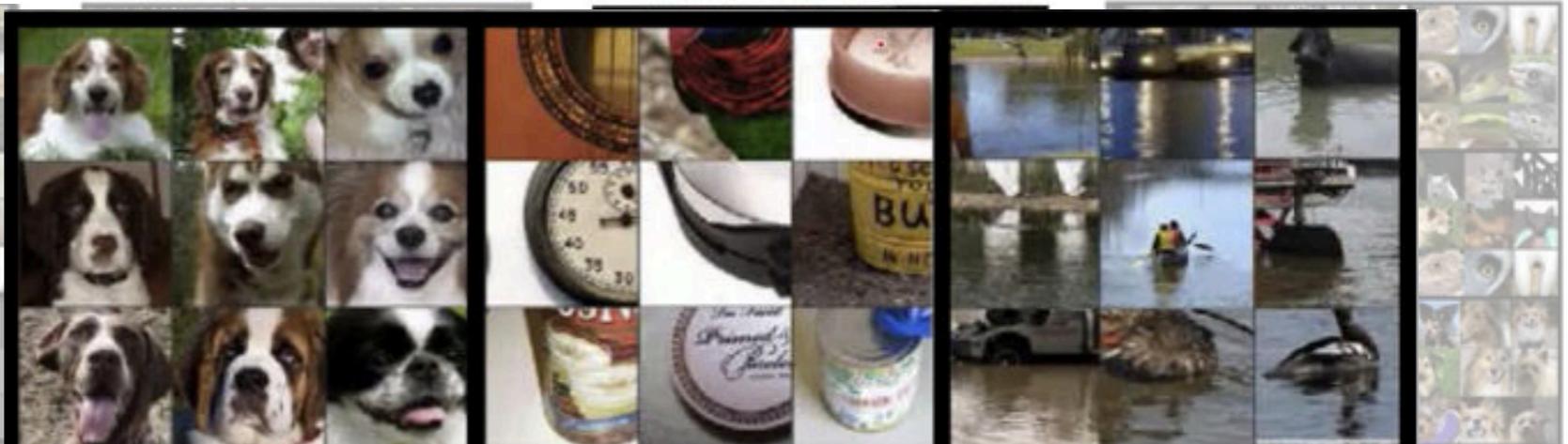
Layer 1



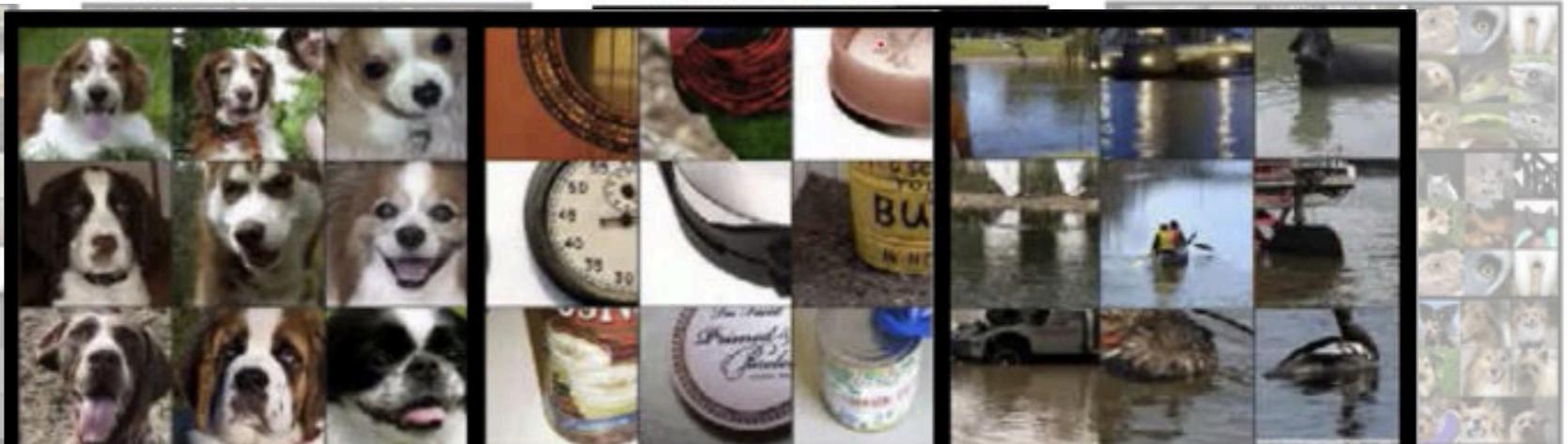
Layer 2



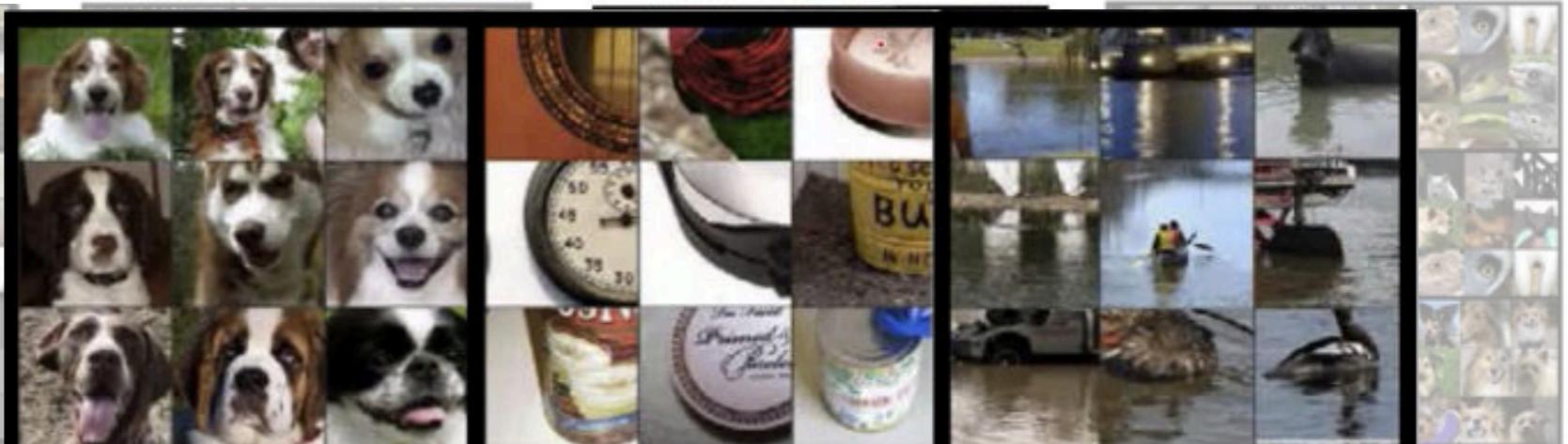
Layer 3



Layer 4



Layer 5



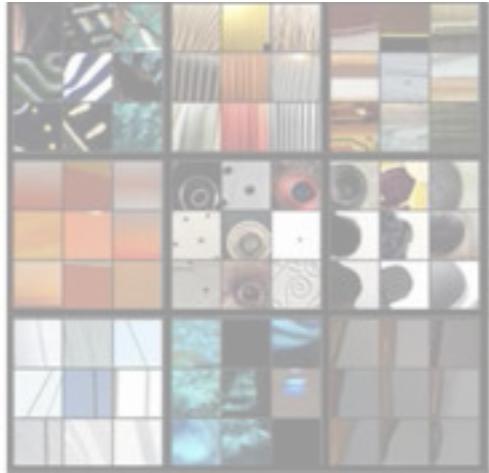
Transferência de estilo

Intuição: o que CNNs aprendem?

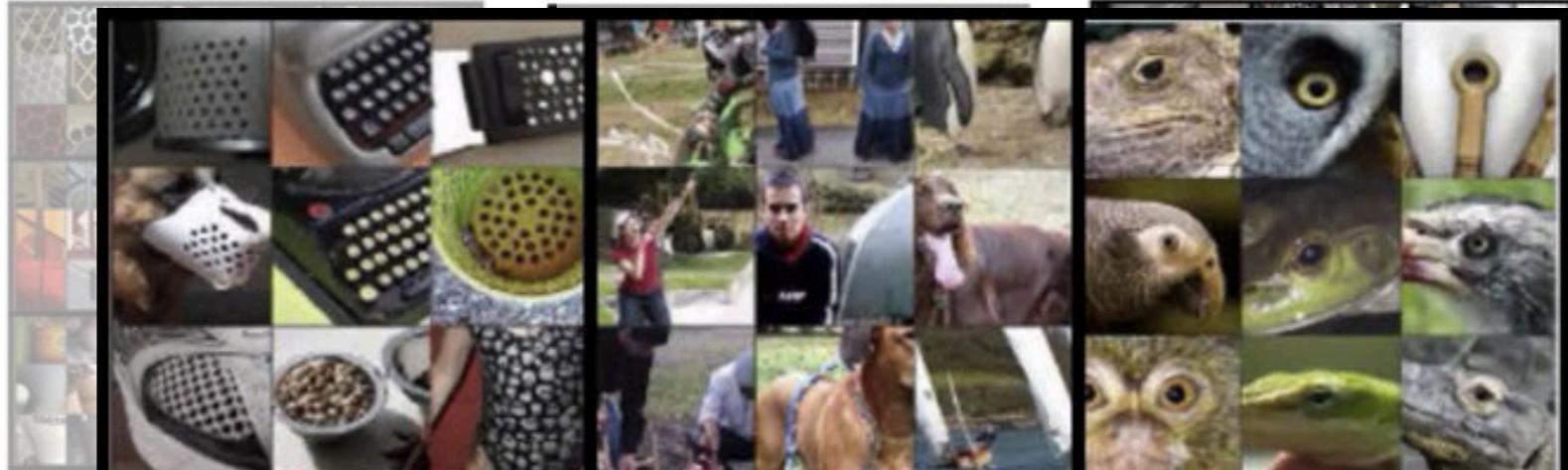
Layer 1



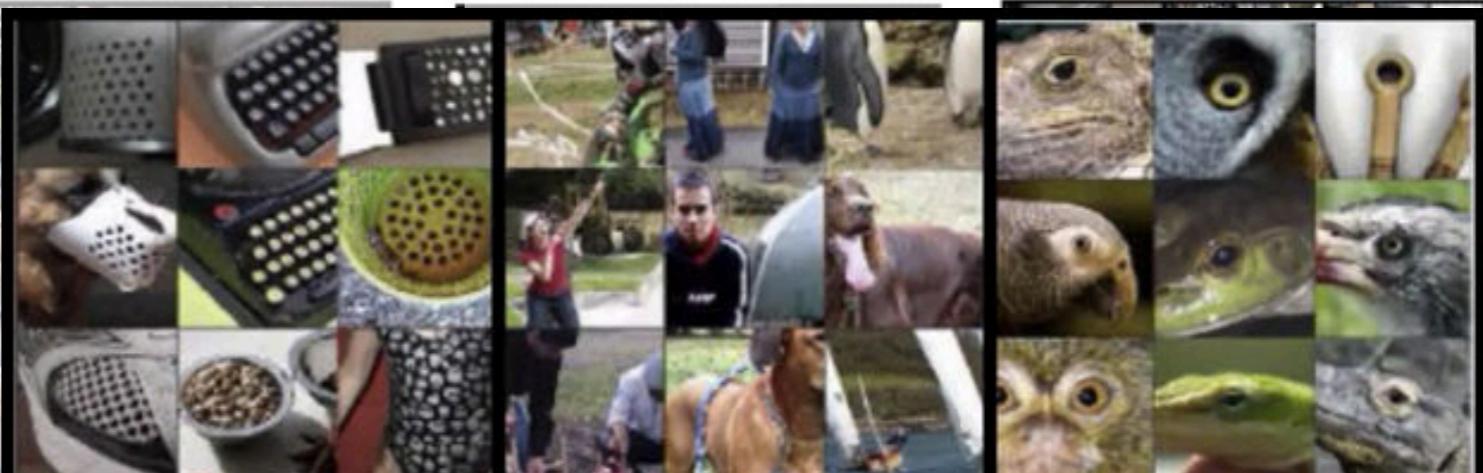
Layer 2



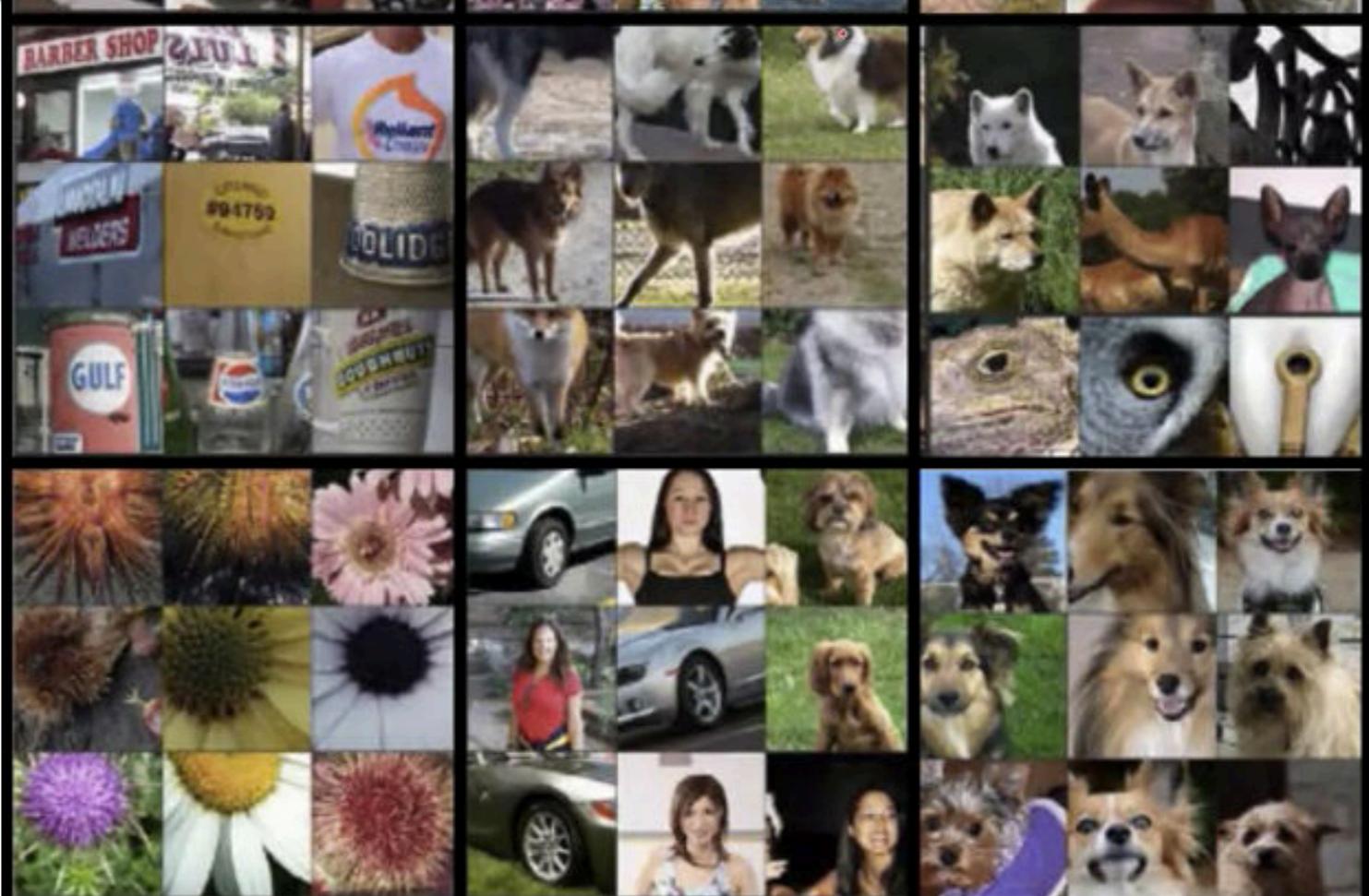
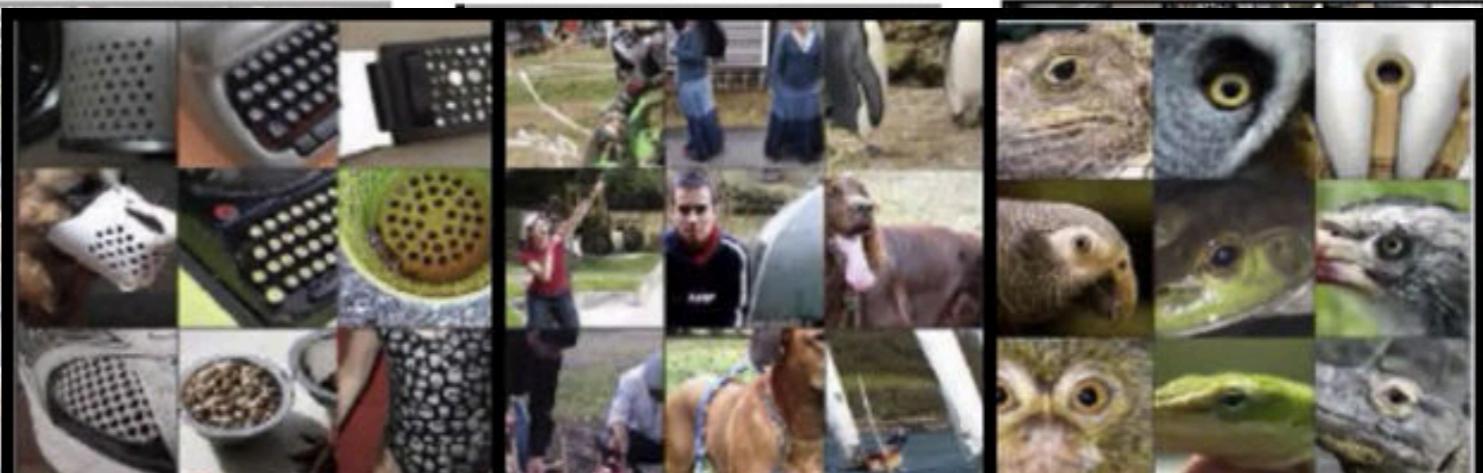
Layer 3



Layer 4

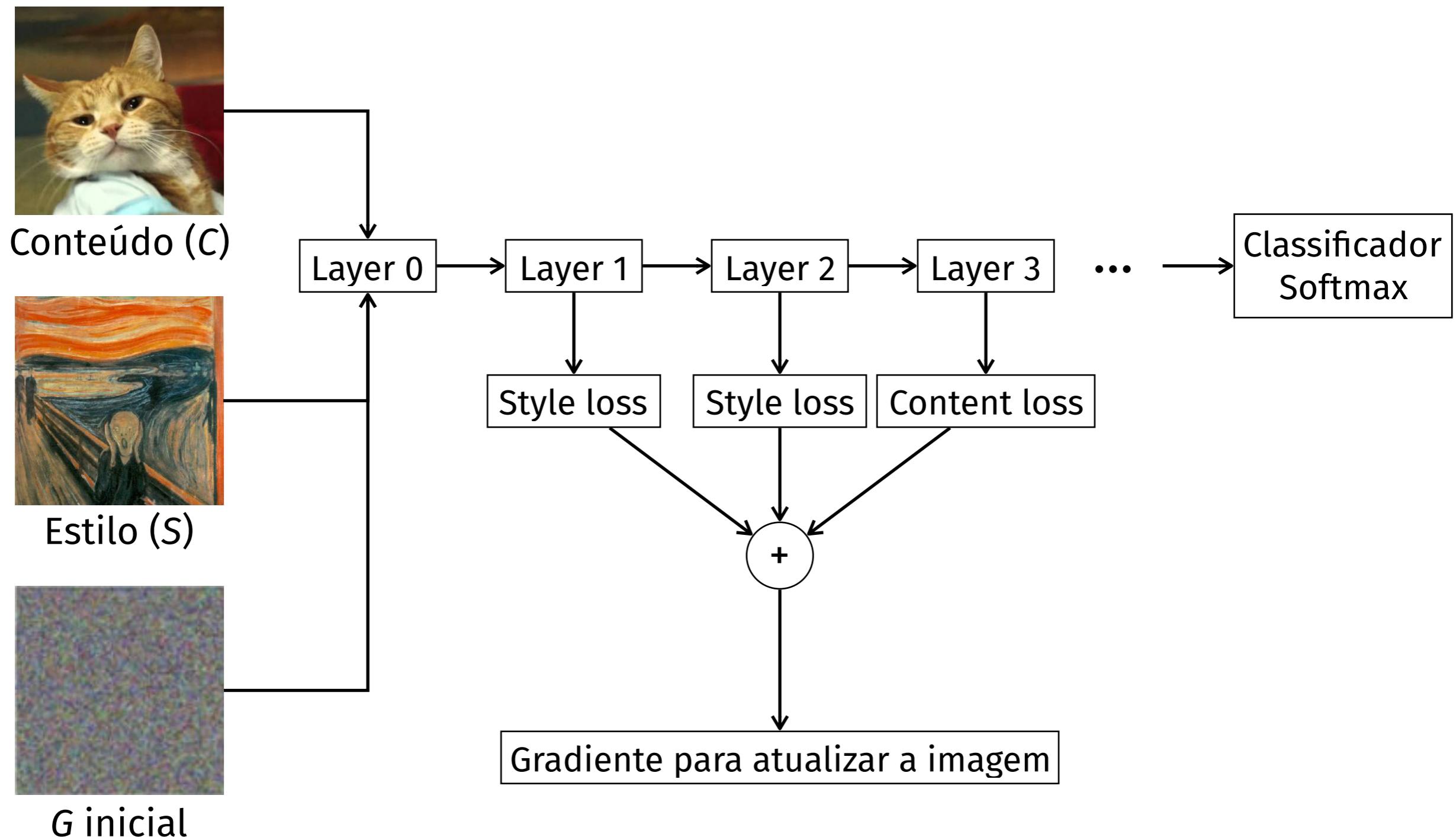


Layer 5



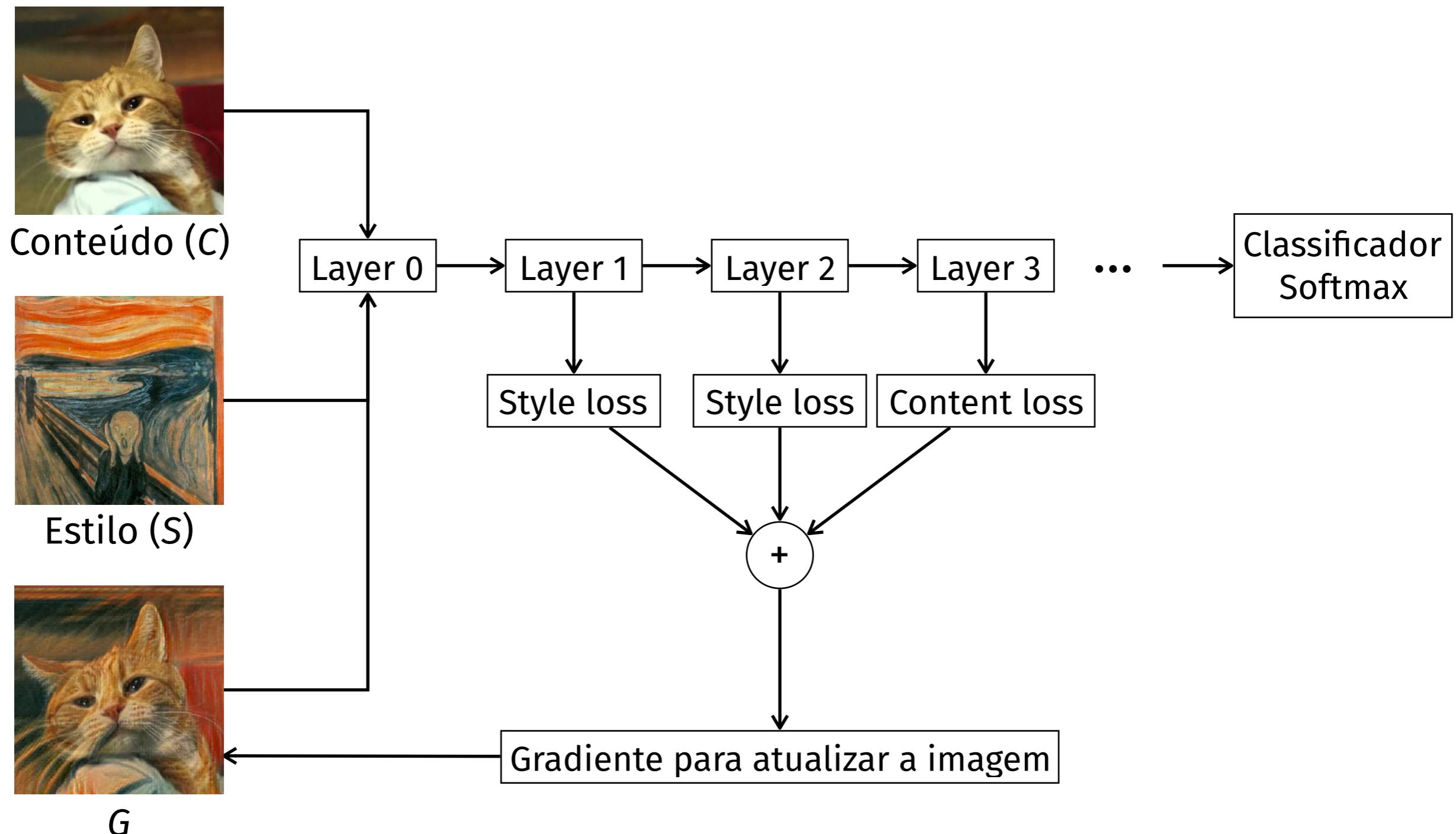
Transferência de estilo

Ideia geral



Transferência de estilo

Ideia geral



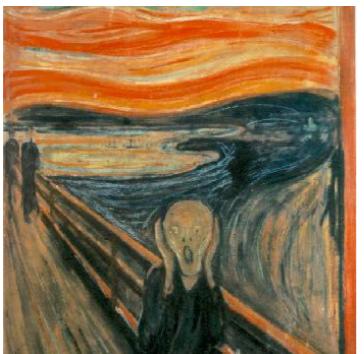
Transferência de estilo

Ideia geral



Como gerar a imagem G ?

Conteúdo (C)



Estilo (S)



Imagen gerada (G)

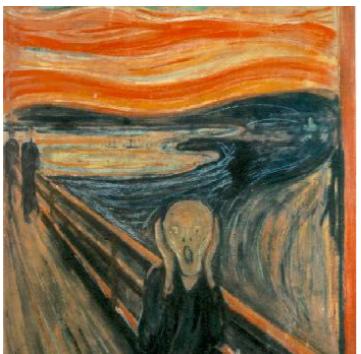
[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Ideia geral



Conteúdo (C)



Estilo (S)



Como gerar a imagem G ?

- Inicializar G randomicamente

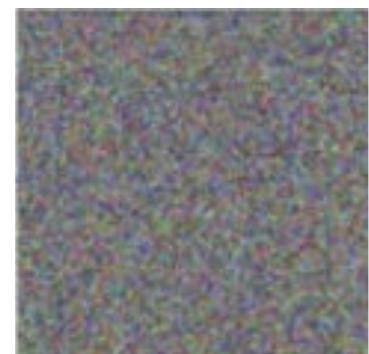
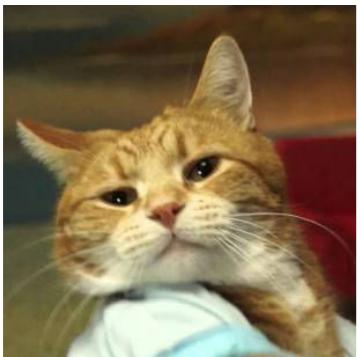


Imagen gerada (G)

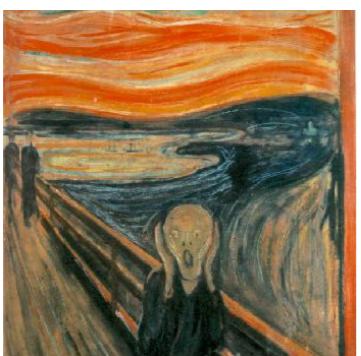
[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Ideia geral



Conteúdo (C)



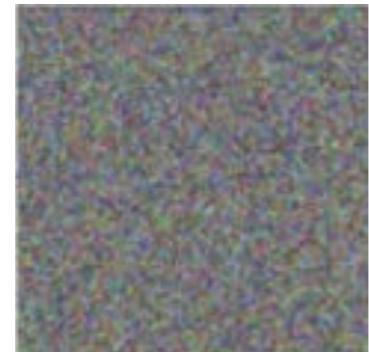
Estilo (S)



Imagen gerada (G)

Como gerar a imagem G ?

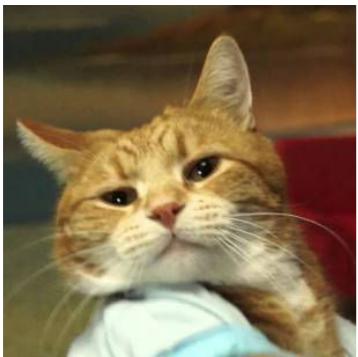
- Inicializar G randomicamente
- Gradient descent para na função de custo



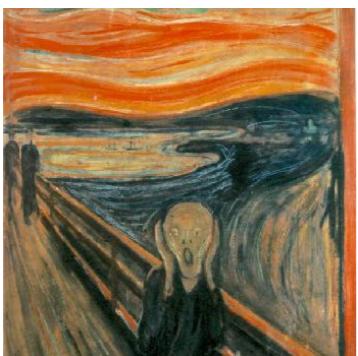
[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Ideia geral



Conteúdo (C)



Estilo (S)



Imagen gerada (G)

Como gerar a imagem G ?

- Inicializar G randomicamente
- Gradient descent para na função de custo
- Atualizar G

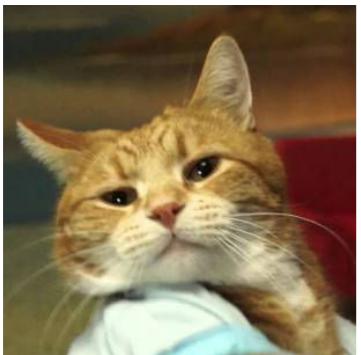
... →



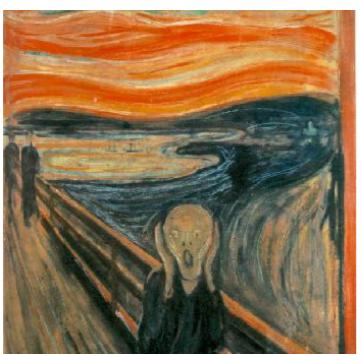
[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Ideia geral



Conteúdo (C)



Estilo (S)



Imagen gerada (G)

Como gerar a imagem G ?

- Inicializar G randomicamente
- Gradient descent para na função de custo
- Atualizar G

... →

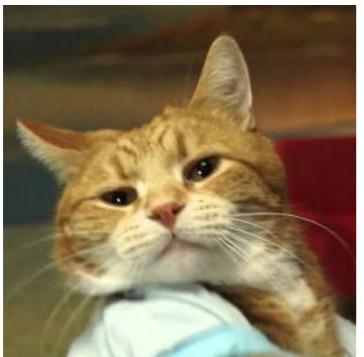


... →

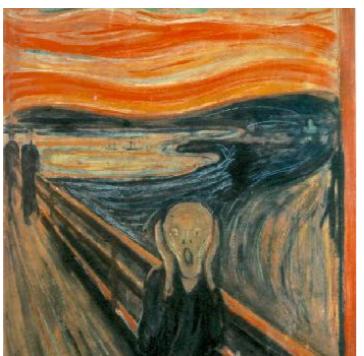


Transferência de estilo

Função de custo



Conteúdo (C)



Estilo (S)



Imagen gerada (G)

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$

[Gatys et al., 2015. A neural algorithm of artistic style]

Função de custo – Content loss

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$

- Escolher CNN pré-treinada (p.ex., VGG)
- Escolher camada l para computar J_{style}
- Sejam $a^{[l]}(C)$ e $a^{[l]}(G)$ as ativações da camada l para as imagens C e G
- Se $a^{[l]}(C)$ e $a^{[l]}(G)$ são similares, então as imagens C e G têm conteúdo similar.

Função de custo – Content loss

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$

- Escolher CNN pré-treinada (p.ex., VGG)
- Escolher camada l para computar J_{style}
- Sejam $a^{[l]}(C)$ e $a^{[l]}(G)$ as ativações da camada l para as imagens C e G
- Se $a^{[l]}(C)$ e $a^{[l]}(G)$ são similares, então as imagens C e G têm conteúdo similar.

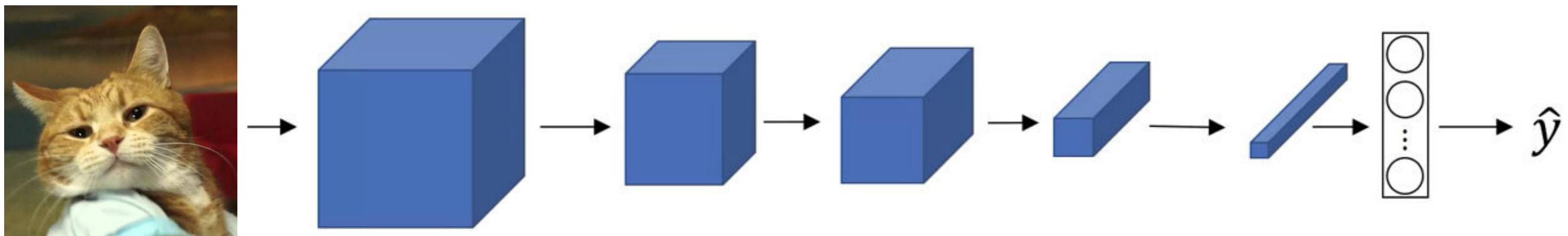
$$J_{content}(C, G) = \|a^{[l]}(C) - a^{[l]}(G)\|^2$$

[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Função de custo – Style loss

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$

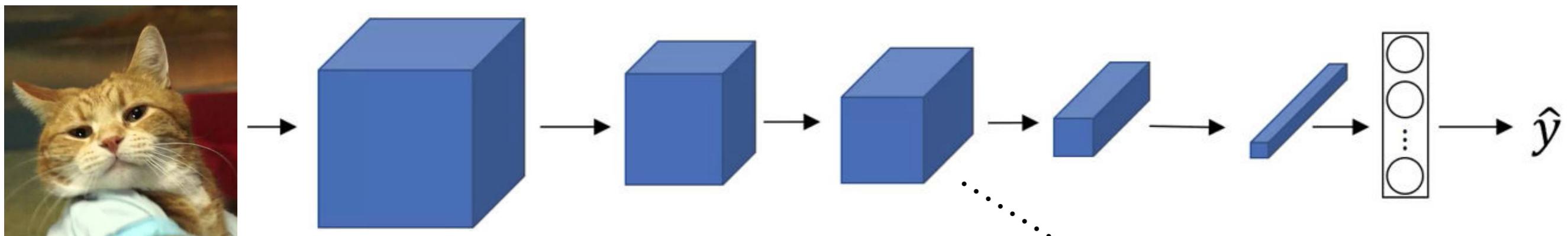


[Gatys et al., 2015. A neural algorithm of artistic style]

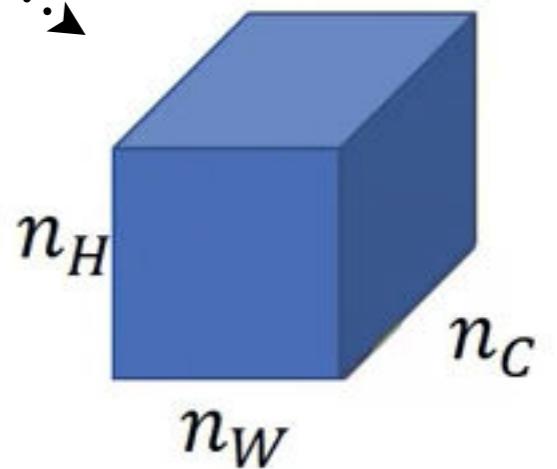
Transferência de estilo

Função de custo – Style loss

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$



- Seja l a camada usada para calcular J_{style} .

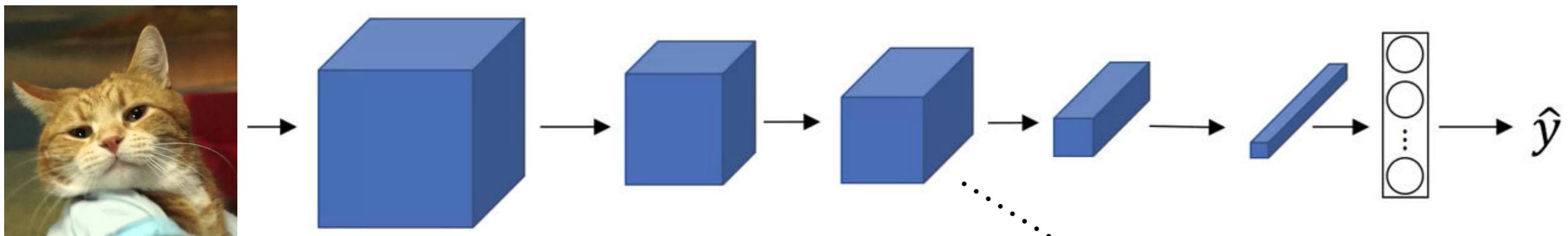


[Gatys et al., 2015. A neural algorithm of artistic style]

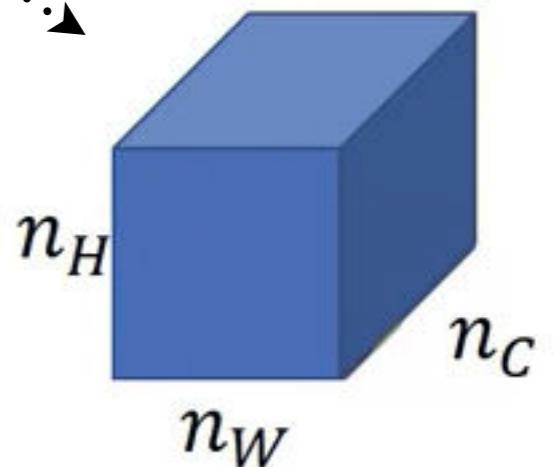
Transferência de estilo

Função de custo – Style loss

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$



- Seja l a camada usada para calcular J_{style} .
- Definir “estilo” como a correlação entre ativações em diferentes canais do output de l .

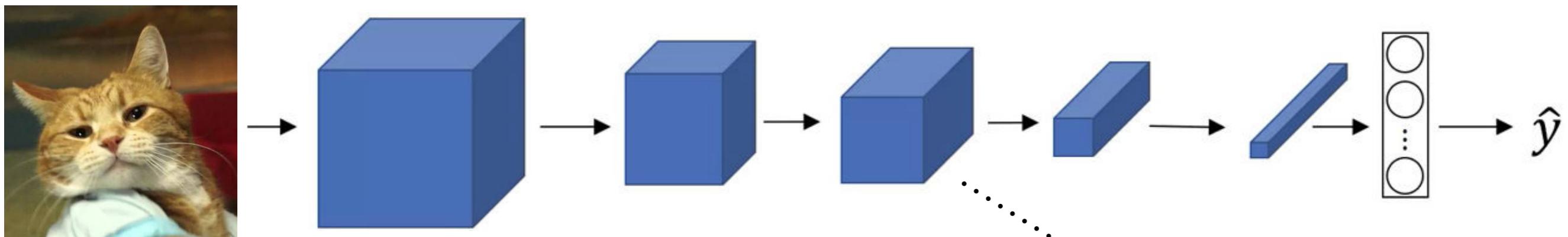


[Gatys et al., 2015. A neural algorithm of artistic style]

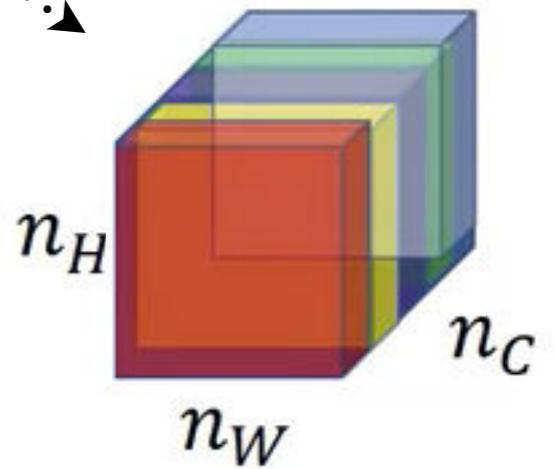
Transferência de estilo

Função de custo – Style loss

$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$



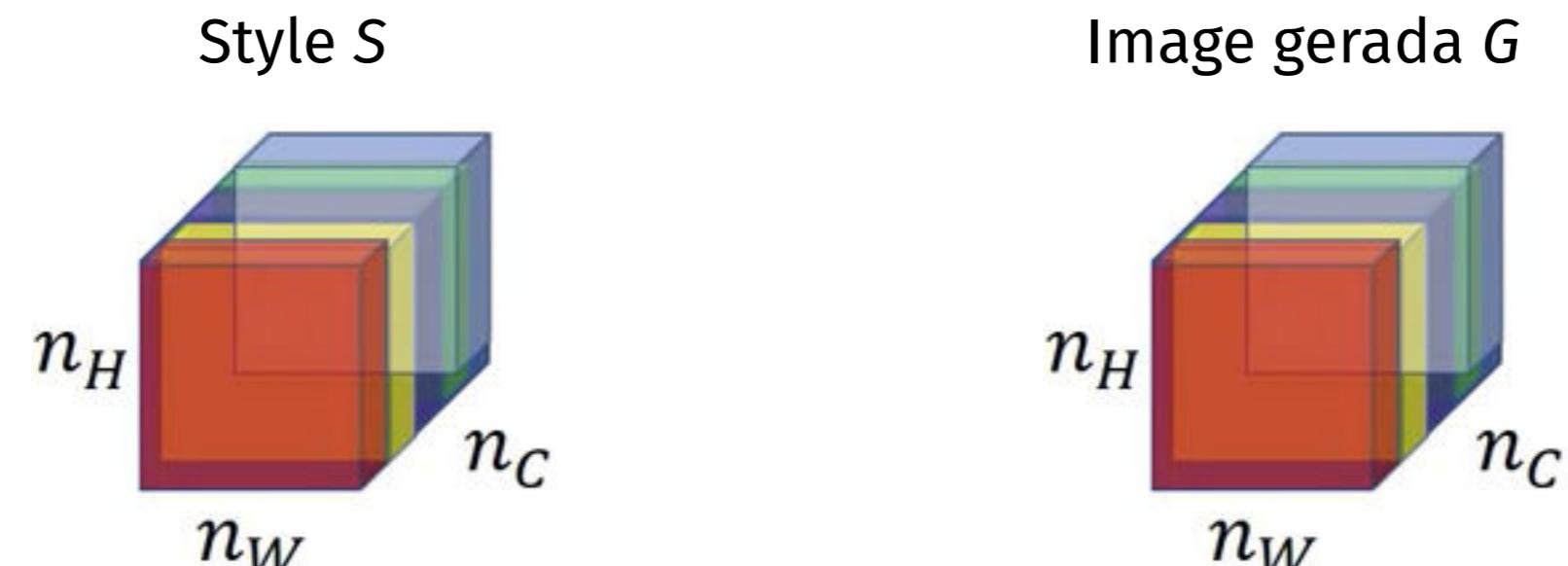
- Seja l a camada usada para calcular J_{style} .
- Definir “estilo” como a correlação entre ativações em diferentes canais do output de l .
- Quão correlacionadas são as ativações dos diferentes canais?



[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

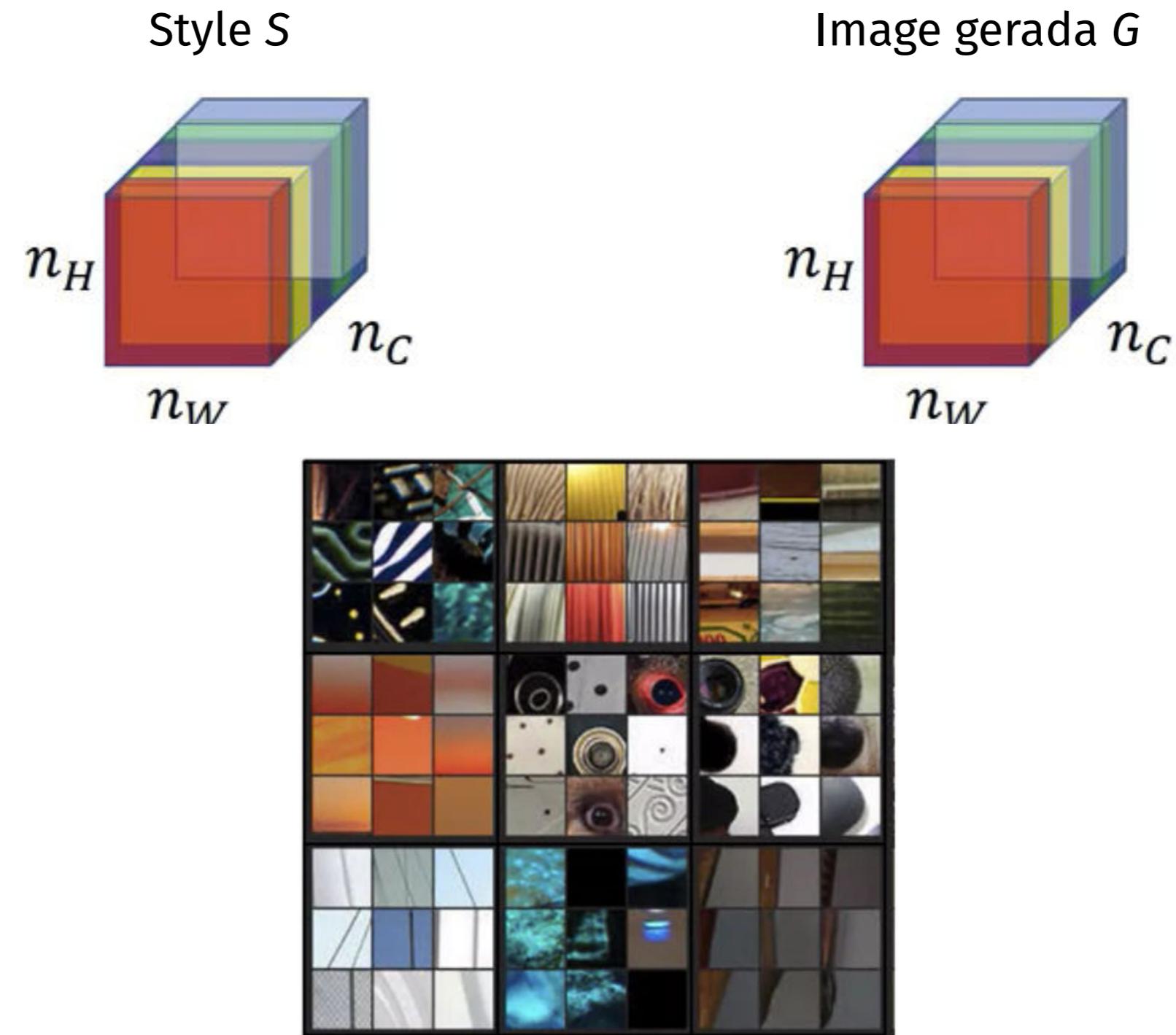
Função de custo – Style loss



[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

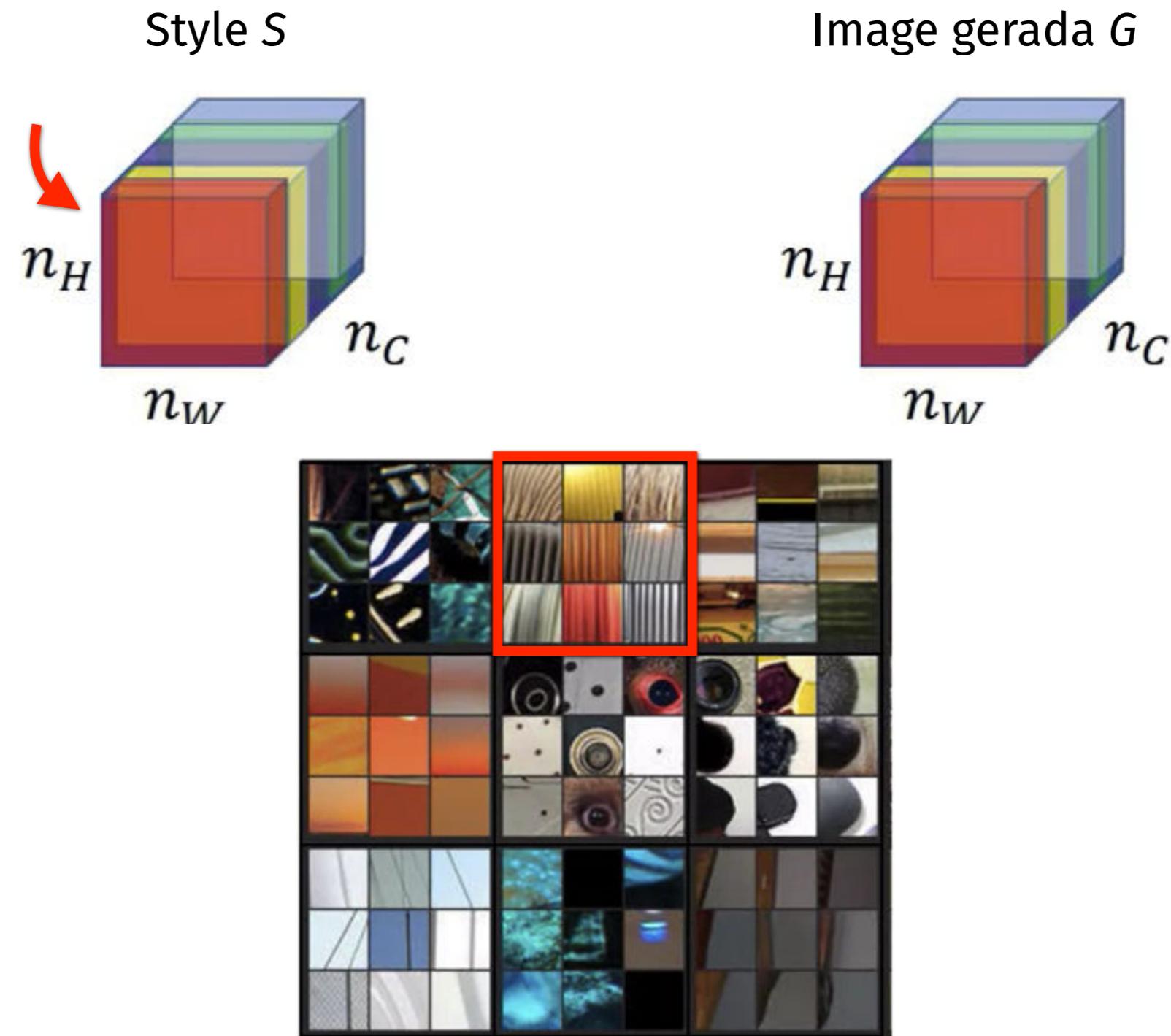
Função de custo – Style loss



[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

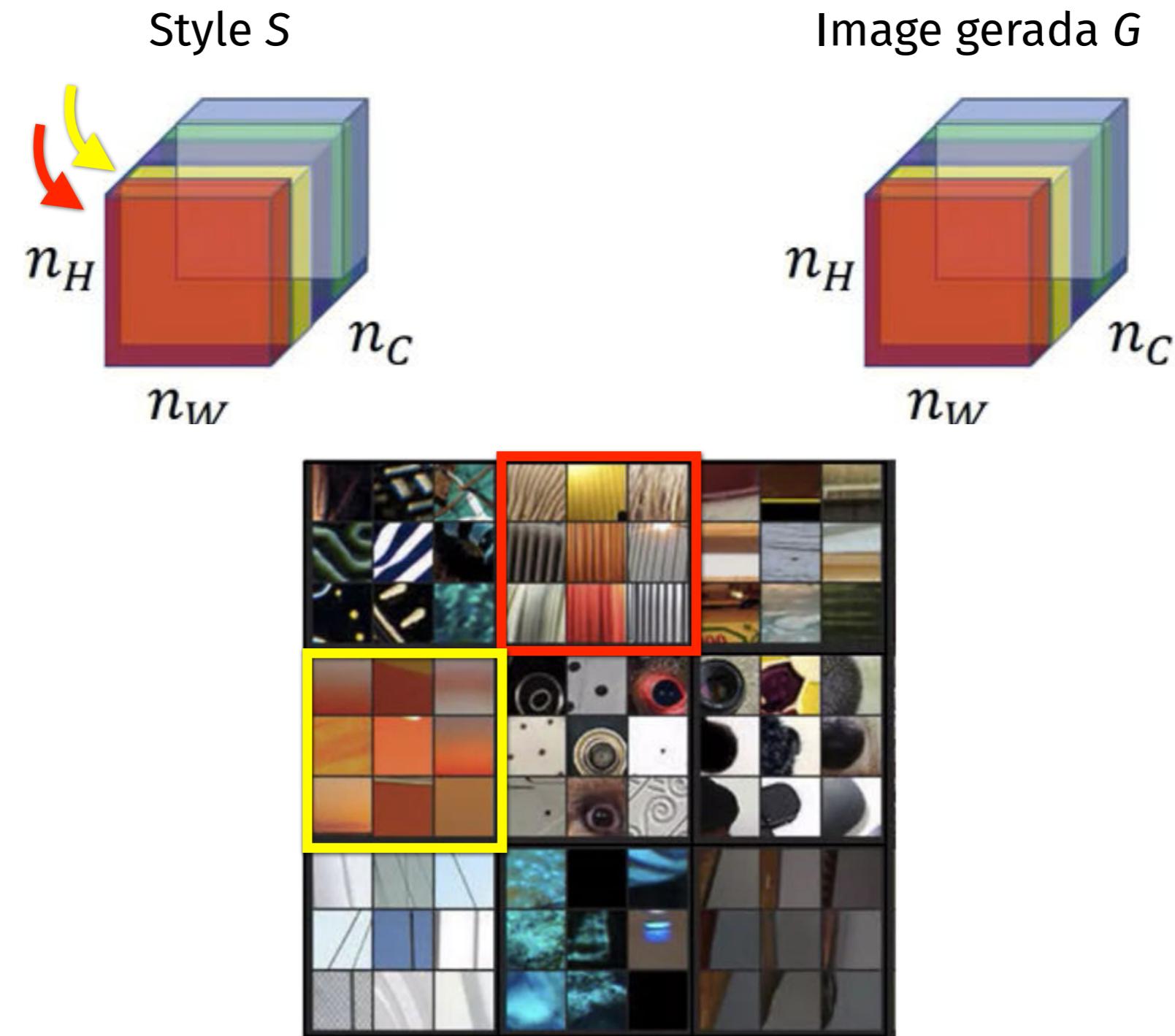
Função de custo – Style loss



[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

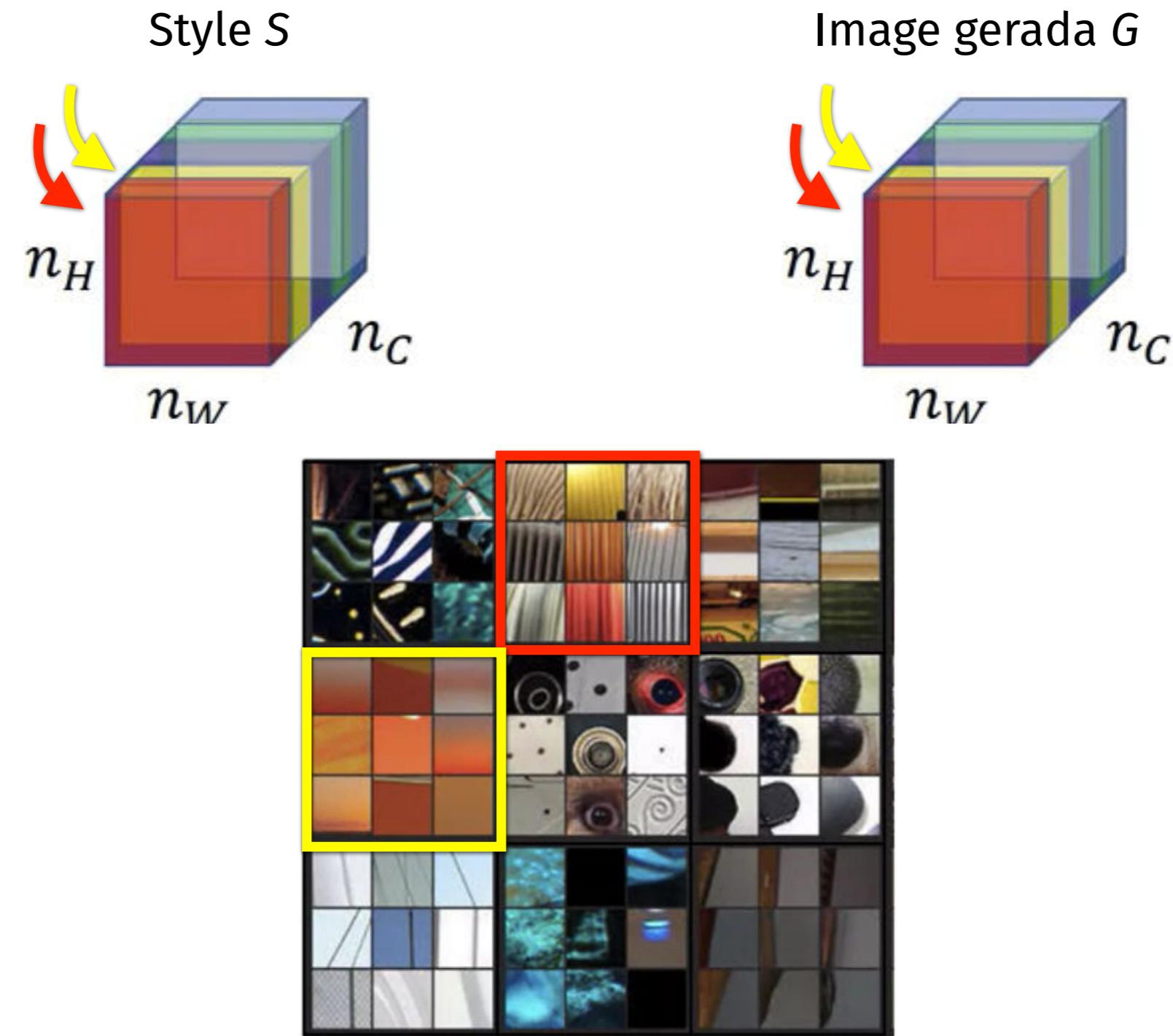
Função de custo – Style loss



[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Função de custo – Style loss

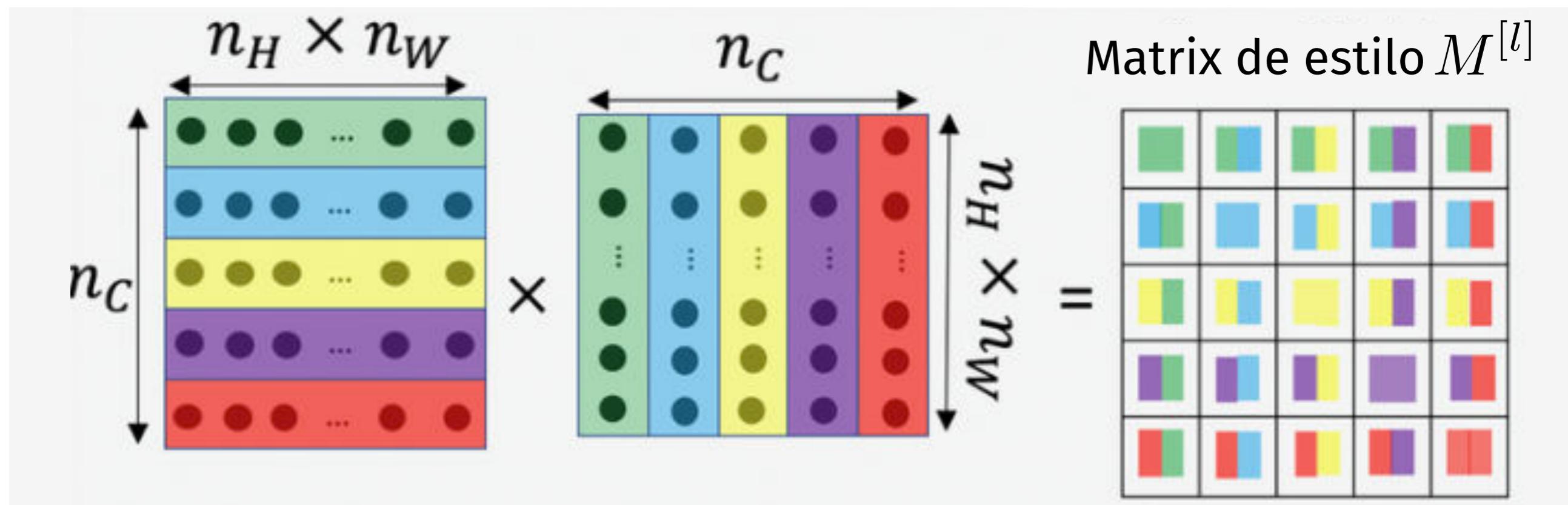


[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Função de custo – Style loss

Matrix de estilo

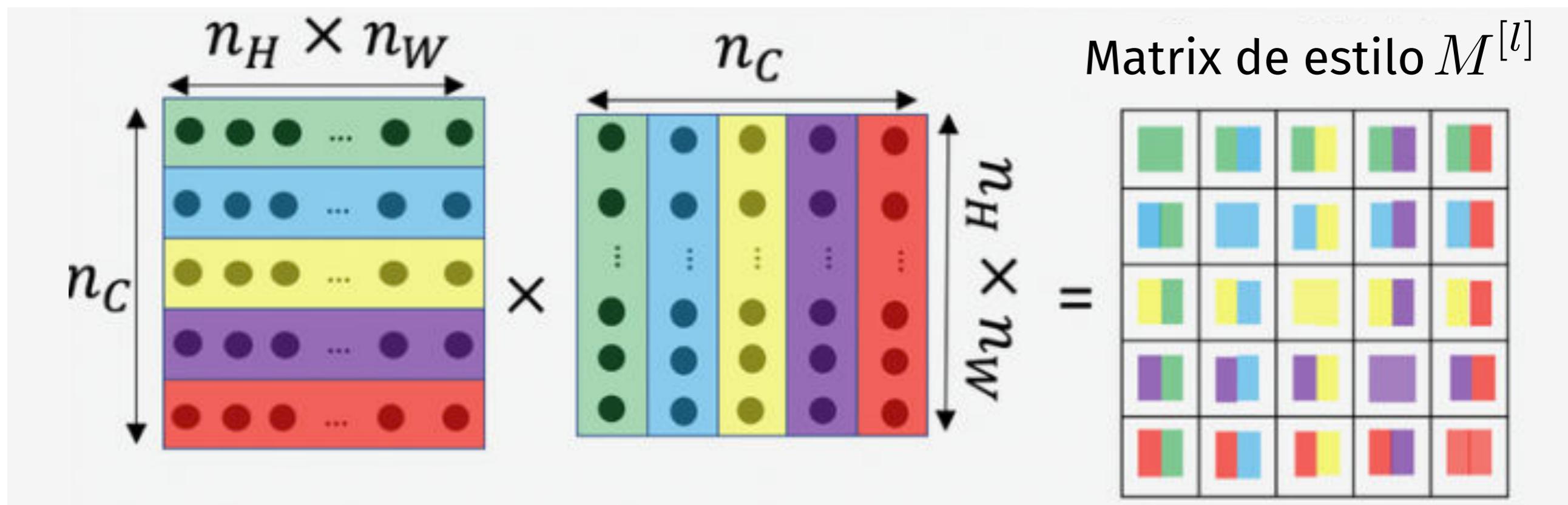


[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Função de custo – Style loss

Matrix de estilo



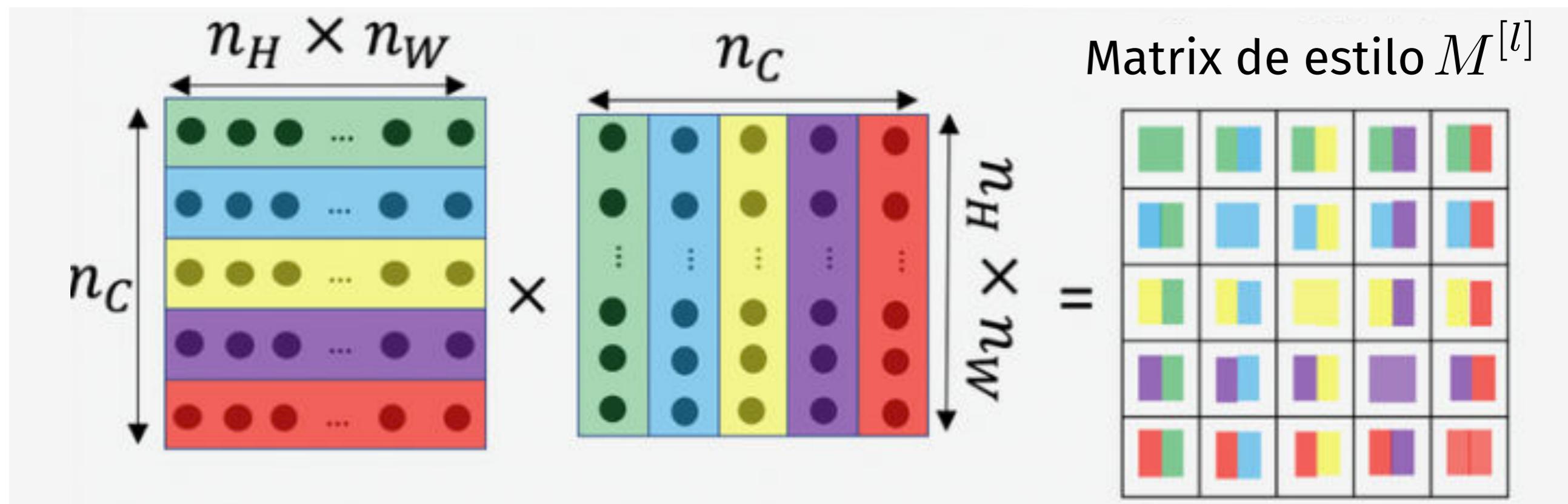
$$J_{style}^{[l]}(S, G) = \|M^{[l]}(S) - M^{[l]}(G)\|^2$$

[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Função de custo – Style loss

Matrix de estilo



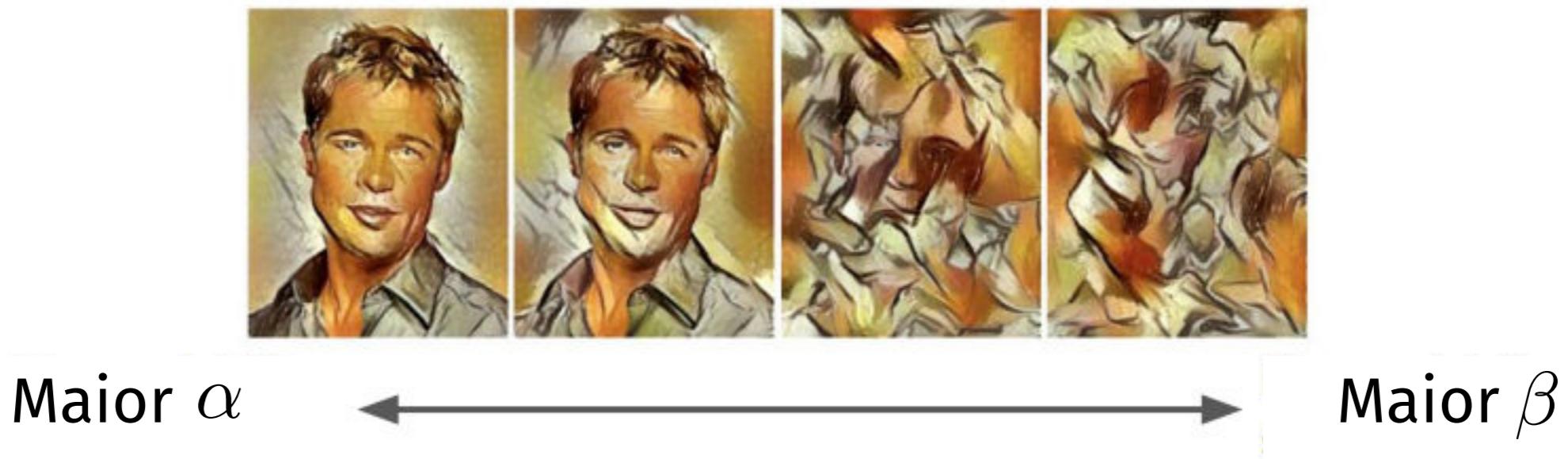
$$J_{style}(S, G) = \sum_l \lambda^{[l]} J_{style}^{[l]}(S, G)$$

[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

Função de custo

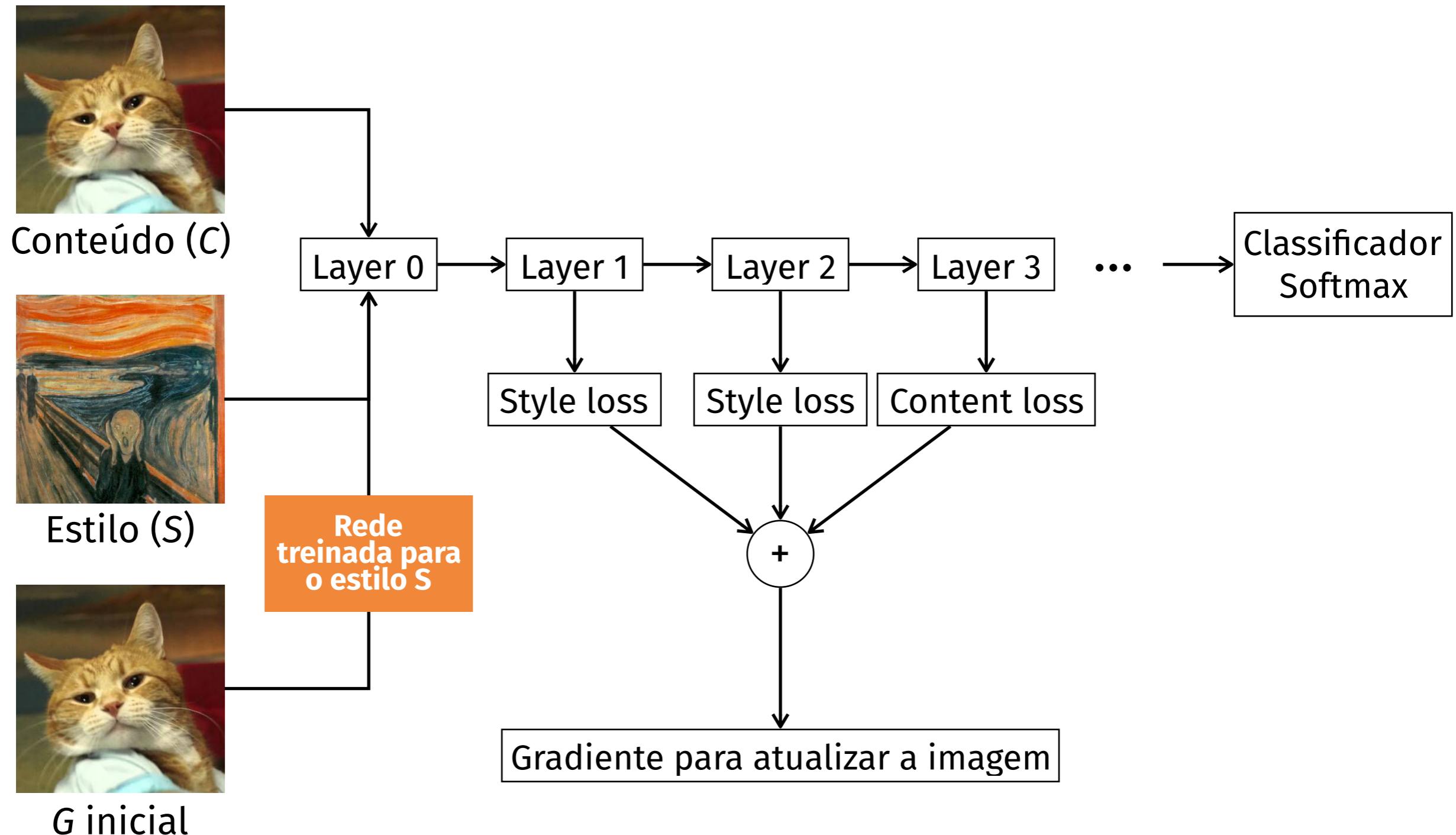
$$J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$$



[Gatys et al., 2015. A neural algorithm of artistic style]

Transferência de estilo

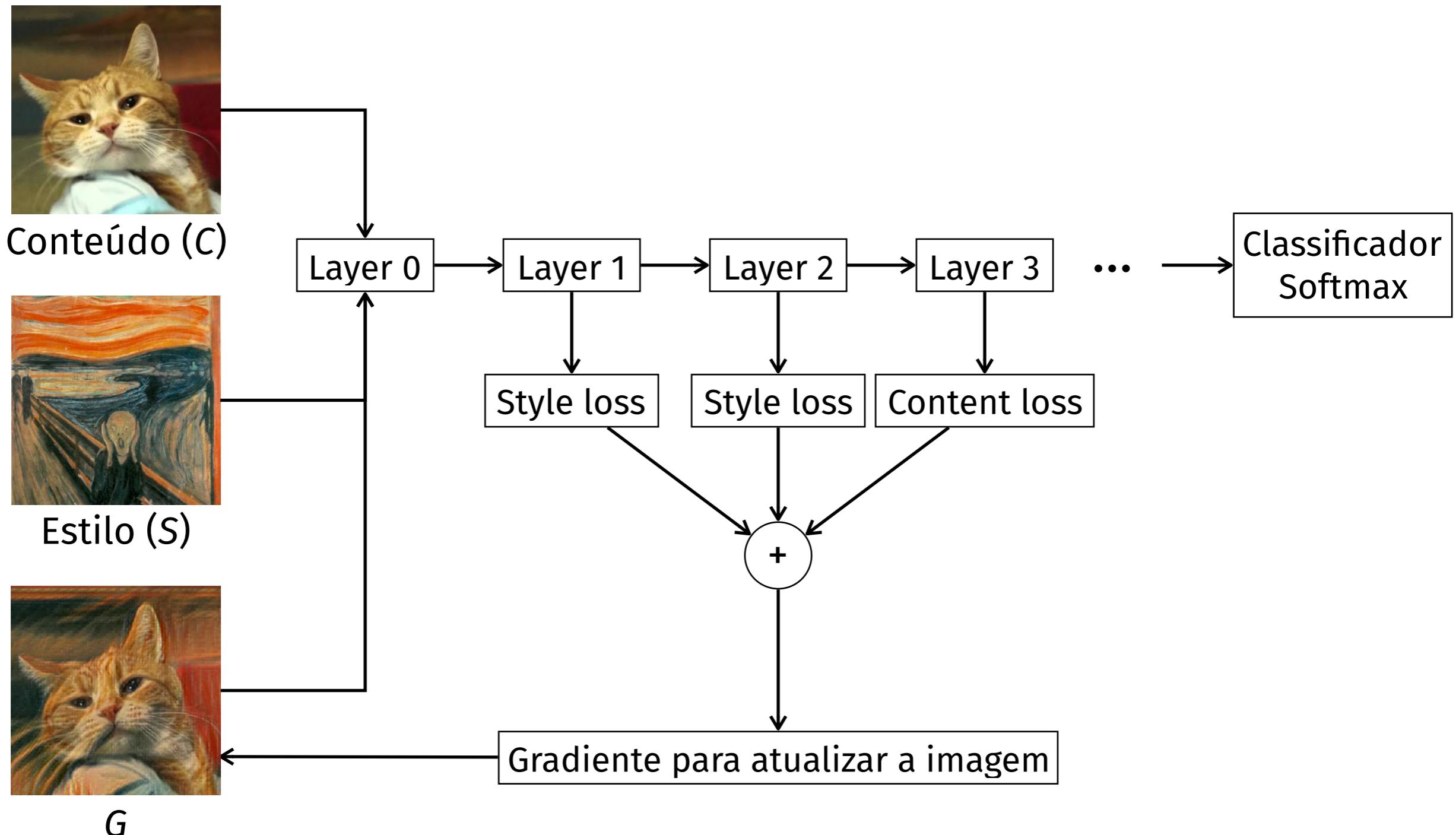
Transferência de estilo rápida



[Johnson et al., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution]

Transferência de estilo

Transferência de estilo rápida



[Johnson et al., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution]

Transferência de estilo

Transferência de estilo rápida



[Johnson et al., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution]

Transferência de estilo

Transferência de estilo rápida

<https://github.com/jcjohnson/fast-neural-style>



[Johnson et al., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution]

Transferência de estilo

Transferência de estilo rápida

<https://github.com/jcjohnson/fast-neural-style>



[Johnson et al., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution]