Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería de Transporte y Logística


# Sistemas Urbanos Inteligentes

## Análisis visual de entornos urbanos

Hans Löbel
Dpto. Ingeniería de Transporte y Logística
Dpto. Ciencia de la Computación

¿Cuál le parece un lugar más bonito?


=

X

# ¿Por qué nos gustaría cuantificar la percepción visual?

- El entorno urbano es percibido principalmente de forma visual.
- Esta percepción puede influenciar su intensidad de uso.
- Puede, por ejemplo, fomentar el uso del transporte público.
- Cuantificar esta percepción a escala nos permitiría identificar lugares candidatos para intervención.

# ¿Cómo medir la percepción a escala?

**Which place looks safer ?**

Which place looks **safer**?

Which place looks **livelier**?

Which place looks **more boring**?

Which place looks **wealthier**?

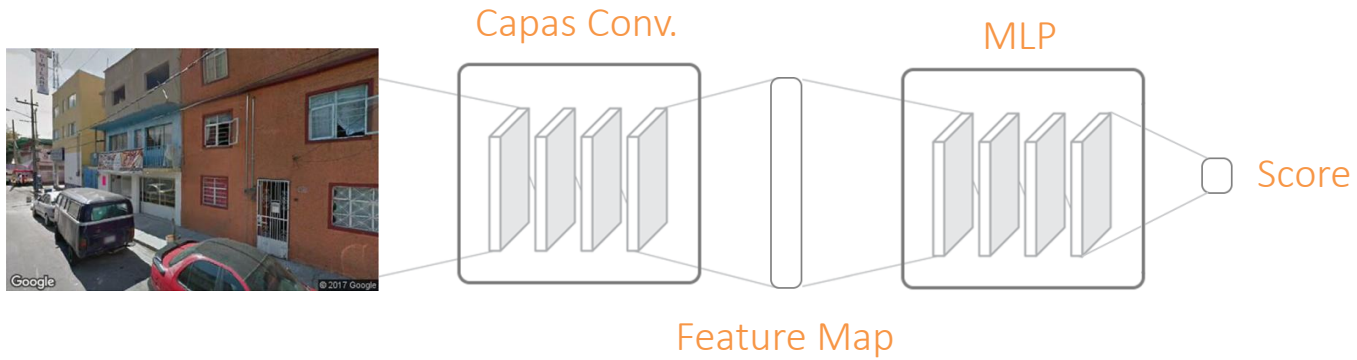
Which place looks **more depressing**?

Which place looks **more beautiful**?



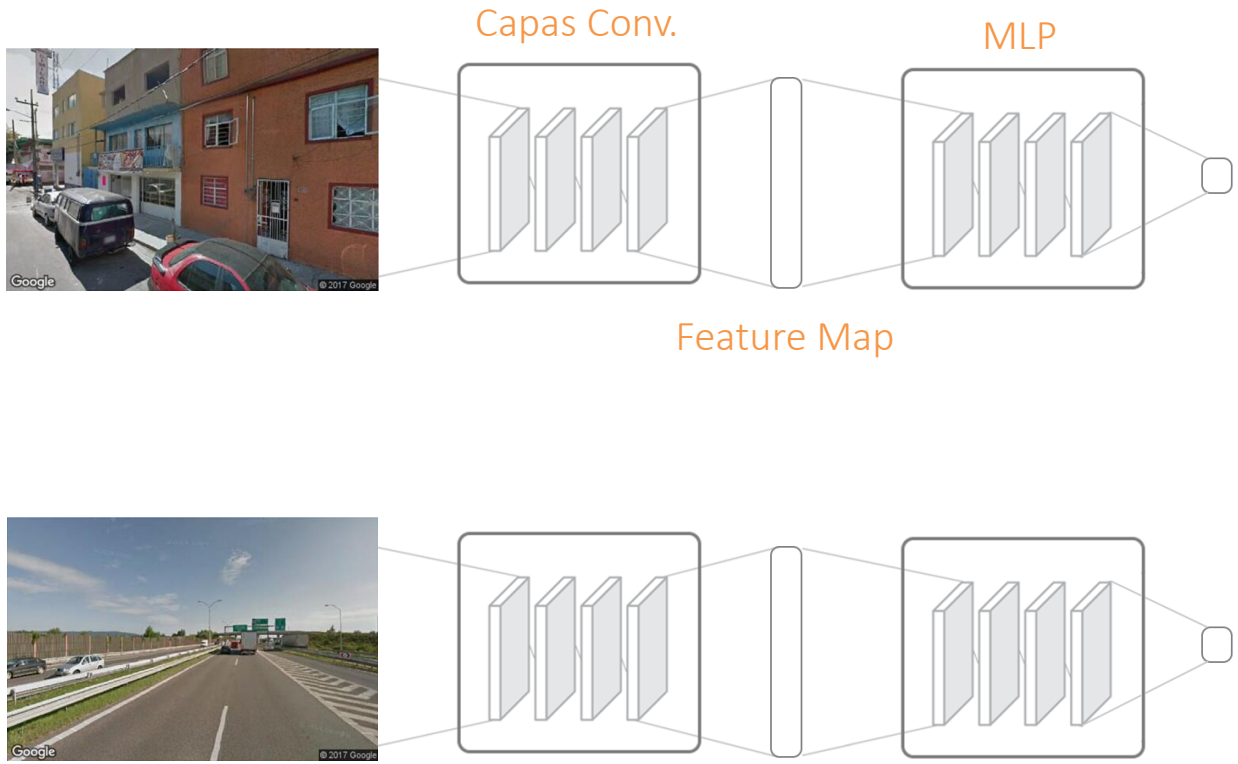
MIT Place Pulse data collection platform - http://pulse.media.mit.edu/

# ¿Cómo medir la percepción a escala?



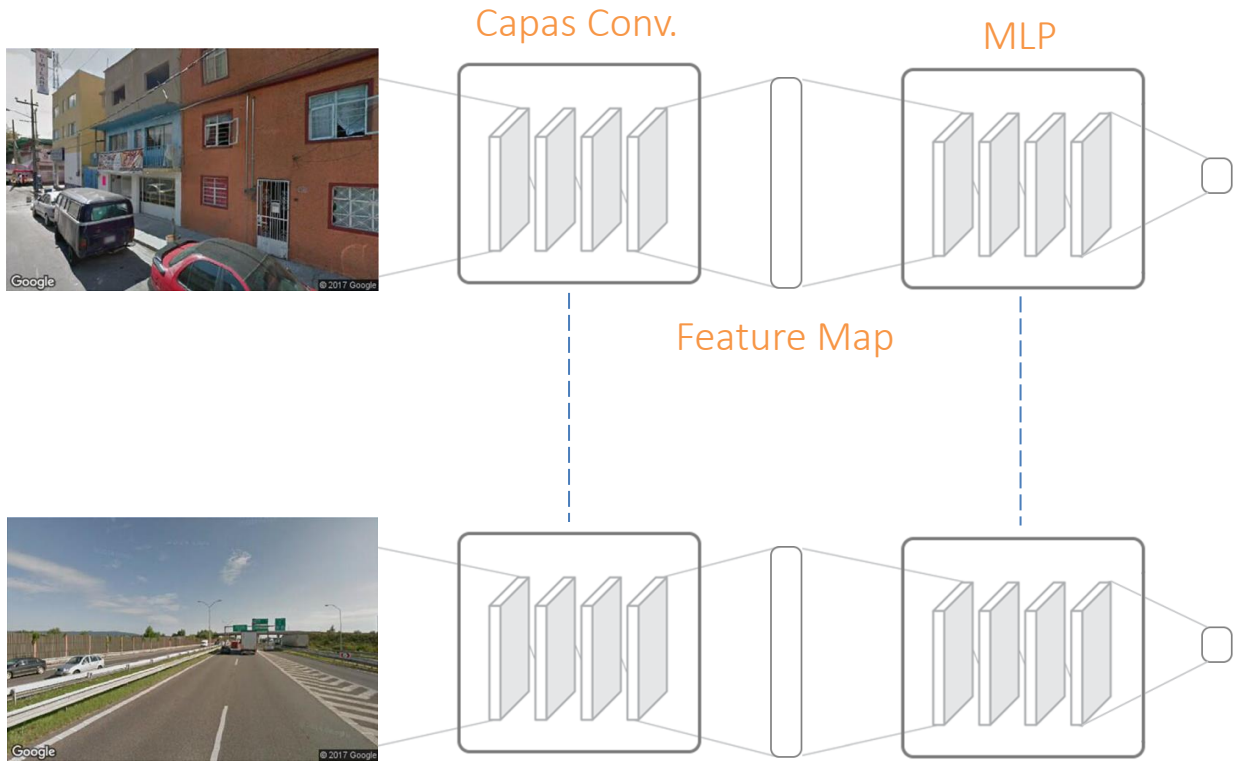

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. European Conference on Computer Vision (ECCV)

# ¿Cómo medir la percepción a escala?



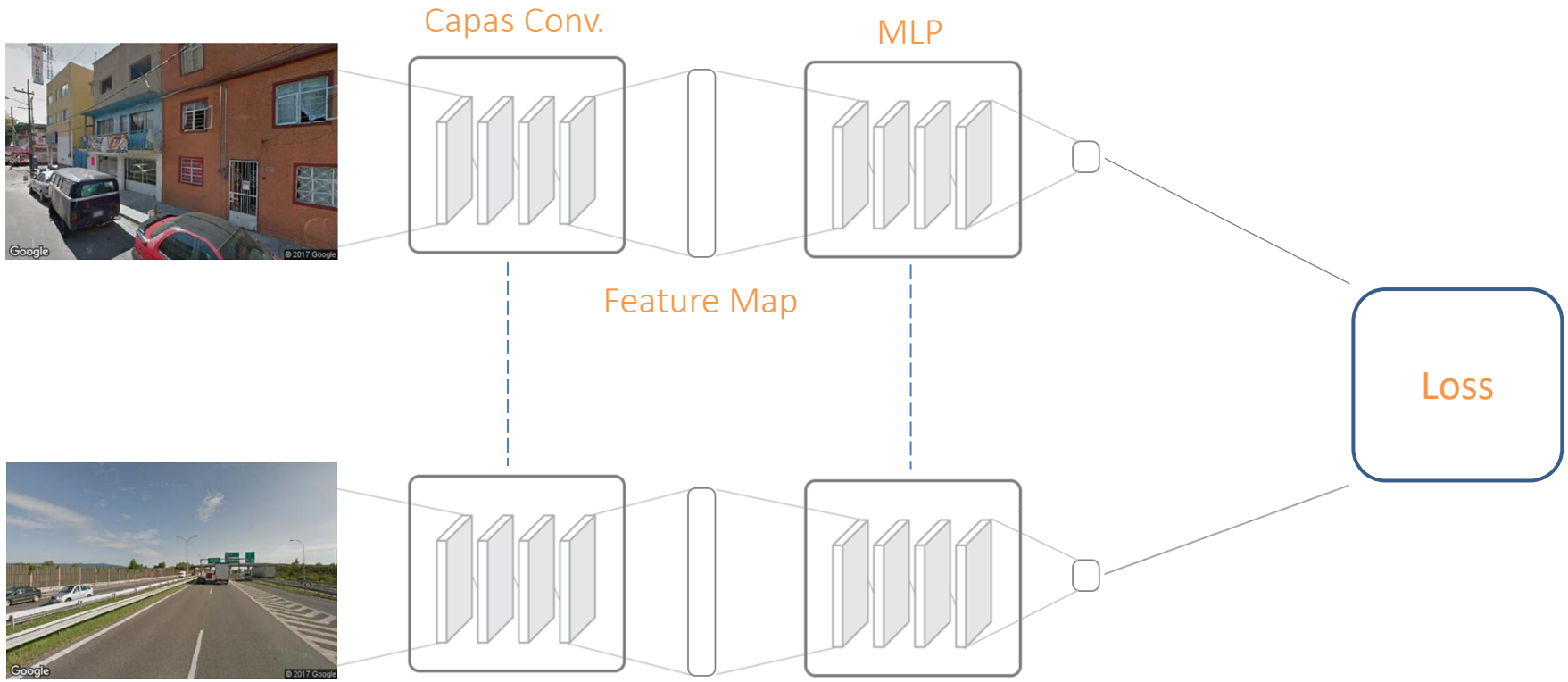

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. European Conference on Computer Vision (ECCV)

# ¿Cómo medir la percepción a escala?



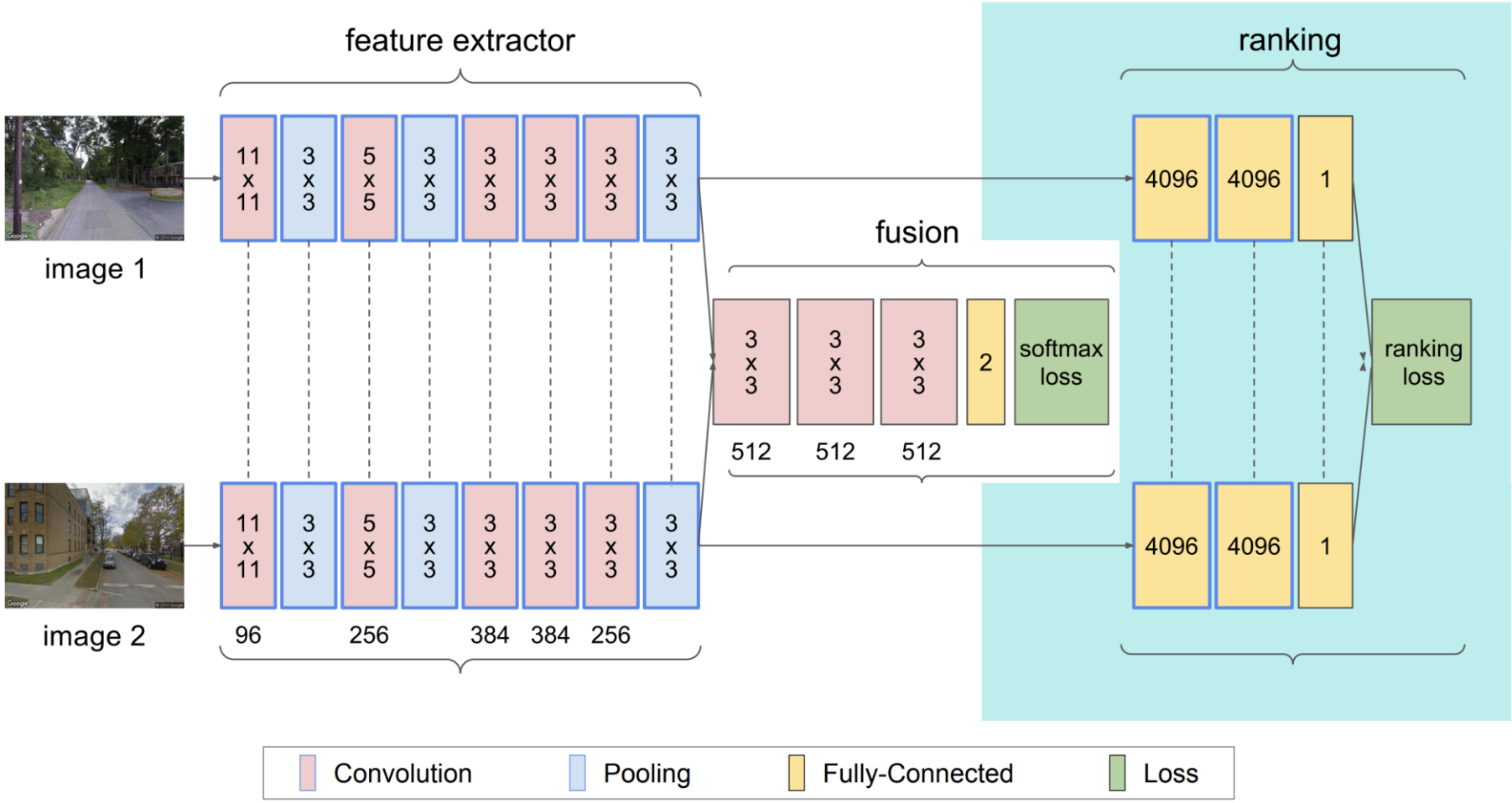

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. European Conference on Computer Vision (ECCV)

# ¿Cómo medir la percepción a escala?




Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. European Conference on Computer Vision (ECCV)

# ¿Cómo medir la percepción a escala?

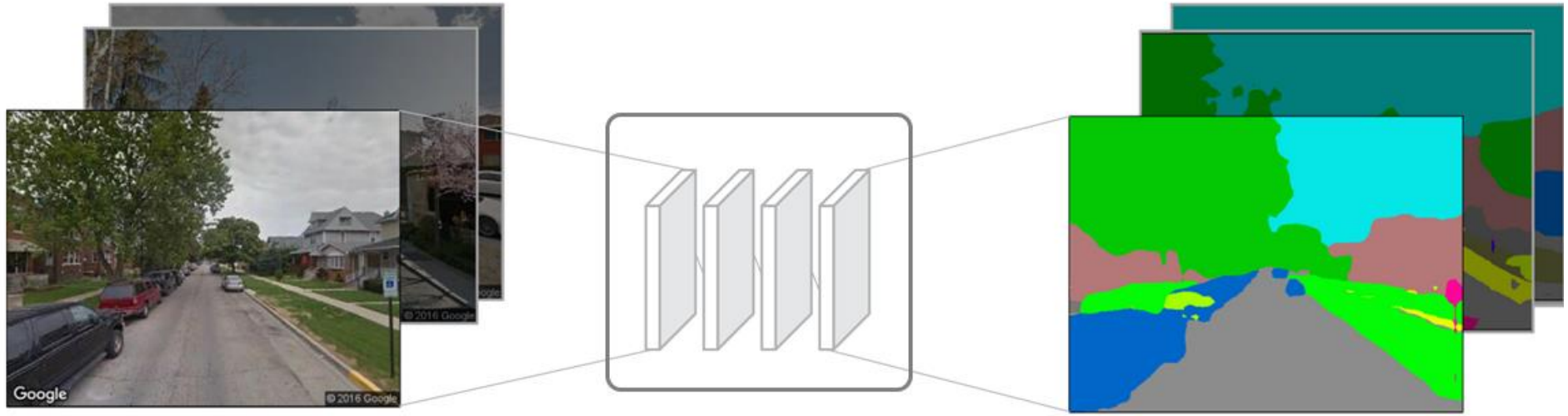



Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. European Conference on Computer Vision (ECCV)

# ¿Cómo medir la percepción a escala?


Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. European Conference on Computer Vision (ECCV)

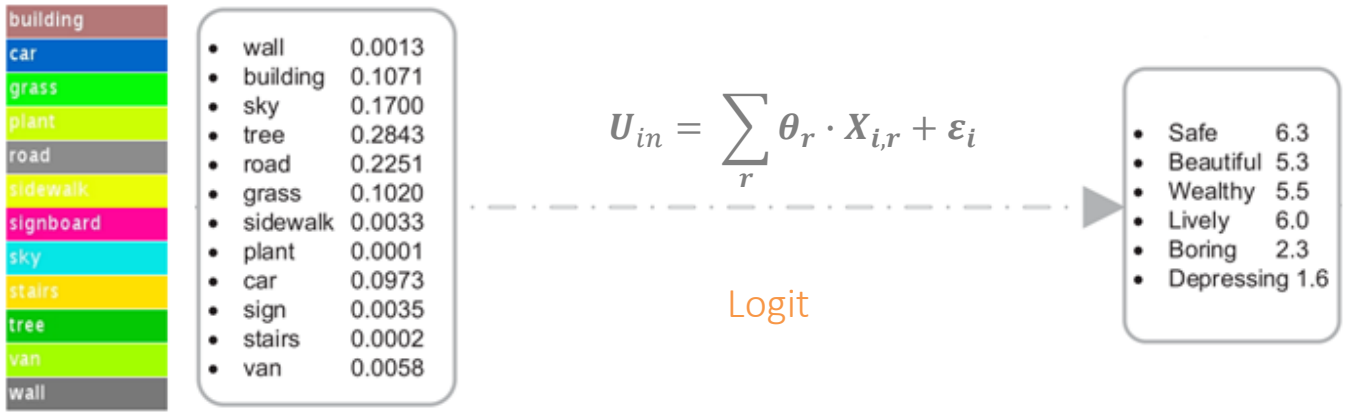
# ¿Cómo explicar la percepción?



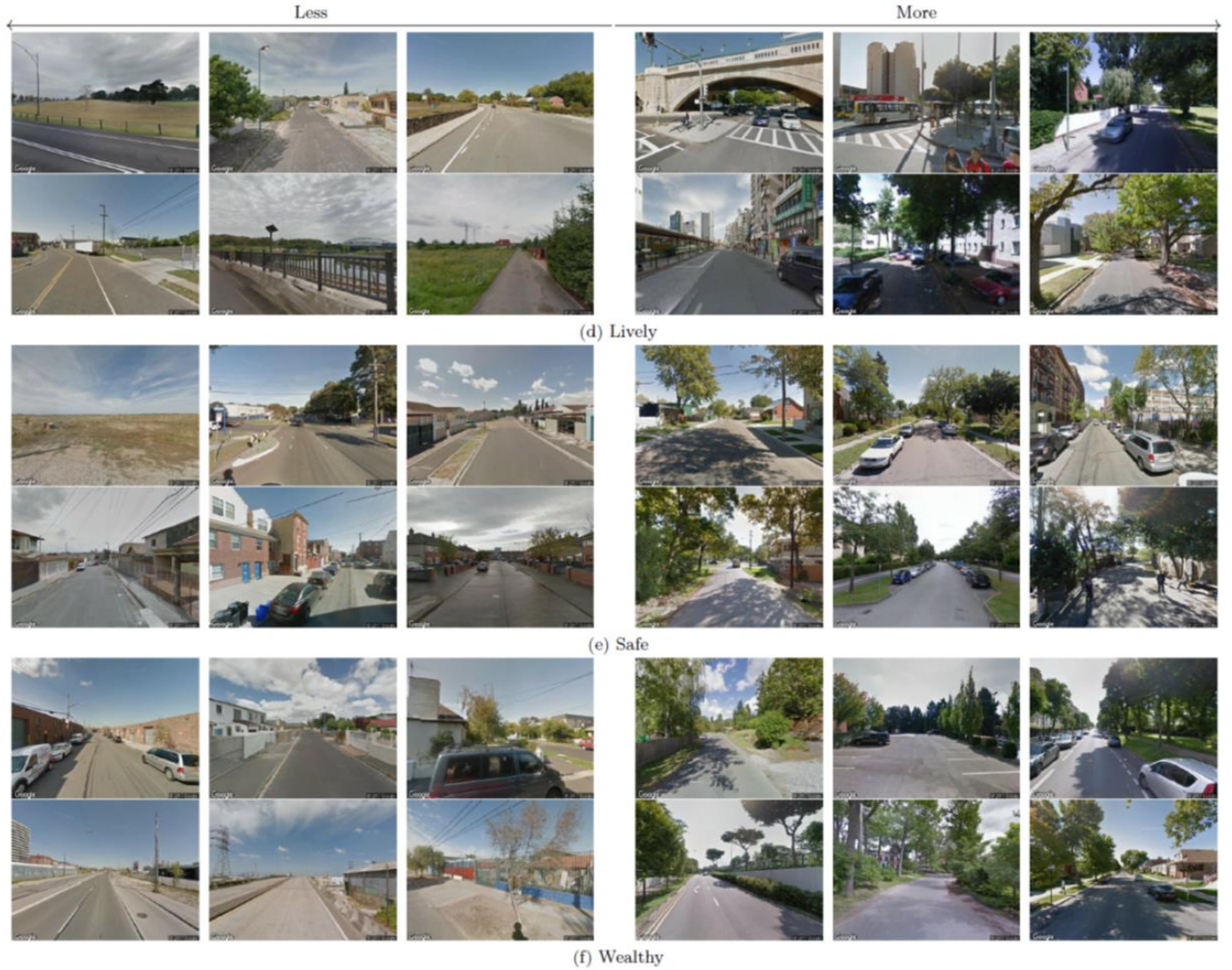

Segmentación semántica de imágenes

Rossetti, T., Lobel, H., Rocco, V., Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. Landscape & Urban Planning, 181, 169-178

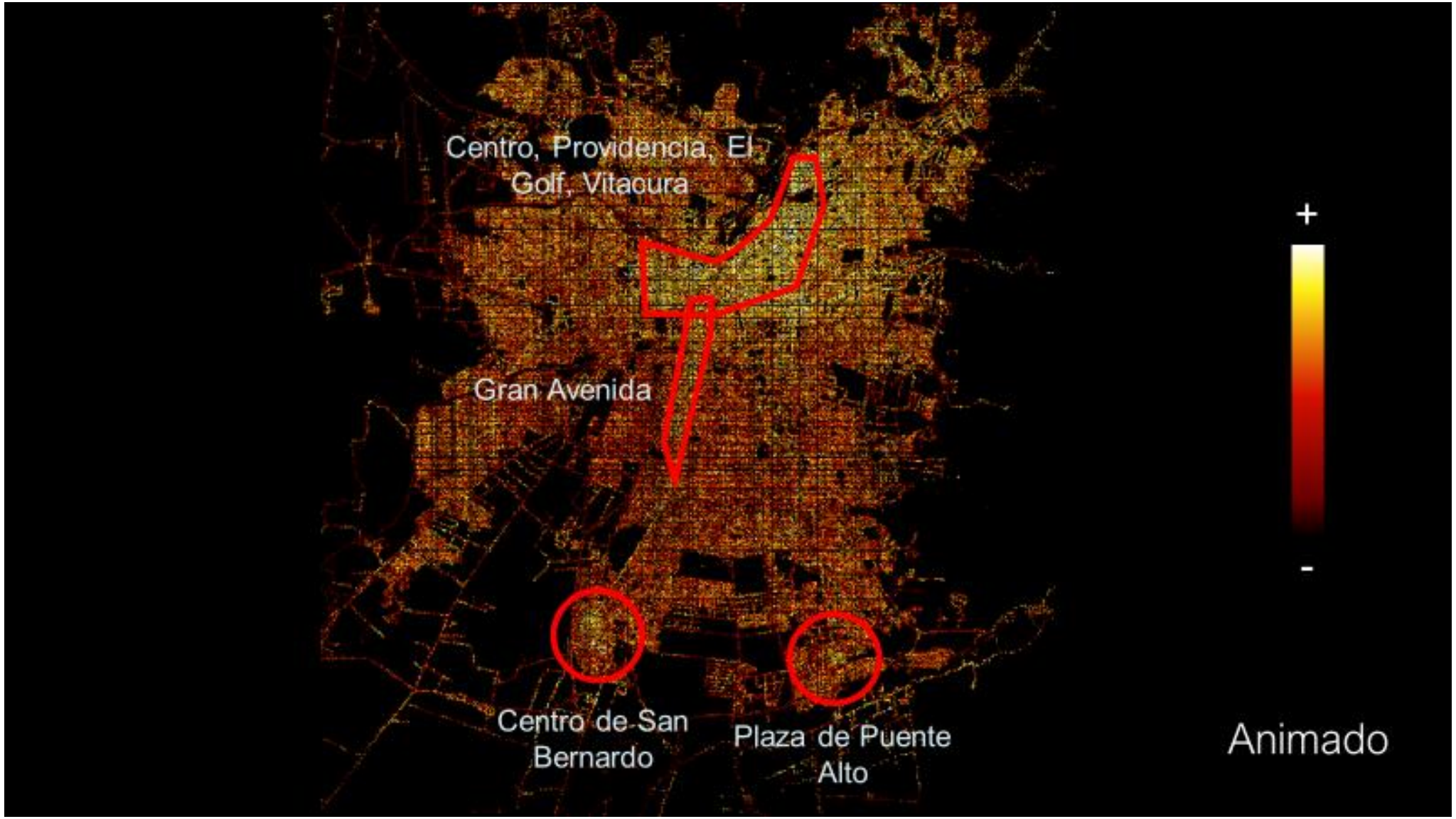
# Atributos ahora son semánticos



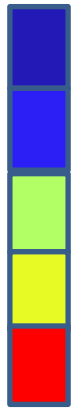

Rossetti, T., Lobel, H., Rocco, V., Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. Landscape & Urban Planning, 181, 169-178




Rossetti, T., Lobel, H., Rocco, V., Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. Landscape & Urban Planning, 181, 169-178

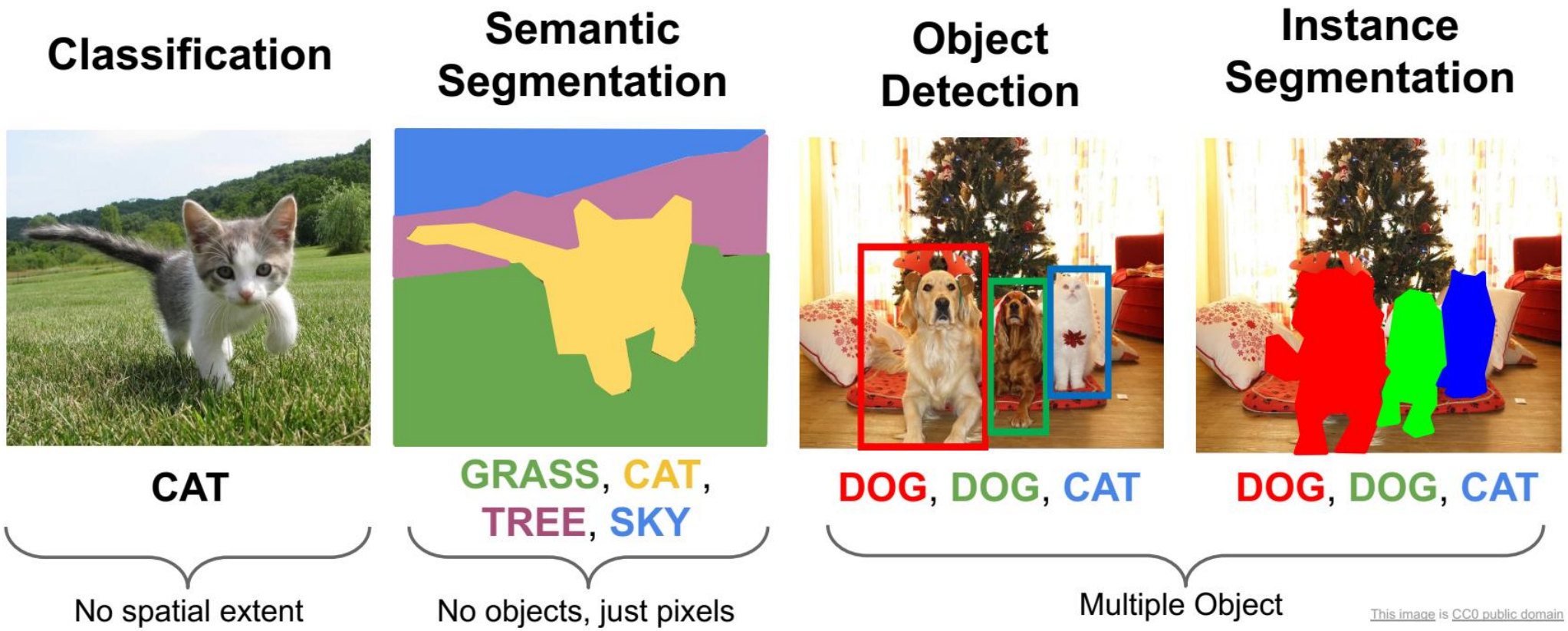

Centro, Providencia, El Golf, Vitacura
Gran Avenida
Centro de San Bernardo
Plaza de Puente Alto
+
-
Animado

Centremos ahora en los cambios con respecto a las CNN que hemos visto


En la segmentación semántica,
cada pixel necesita ser clasificado

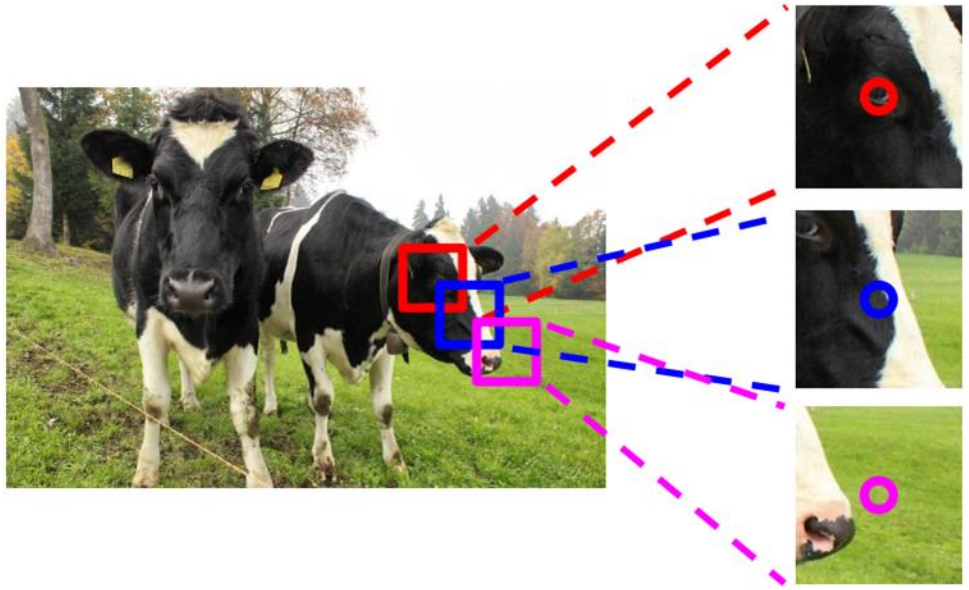
La segmentación semántica es una tarea clásica en visión por computador

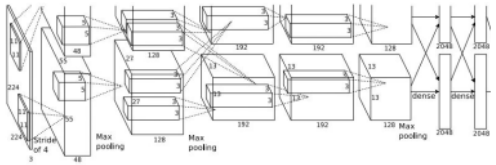


# ¿Por qué usar CNN para segmentación semántica?

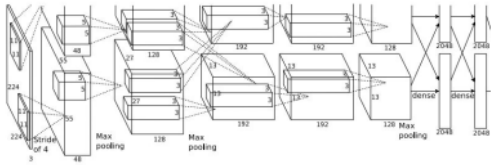

Imposible de clasificar sin algún tipo de contexto

# ¿Cómo usar CNN para segmentación semántica?

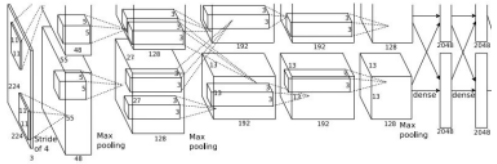




Vaca



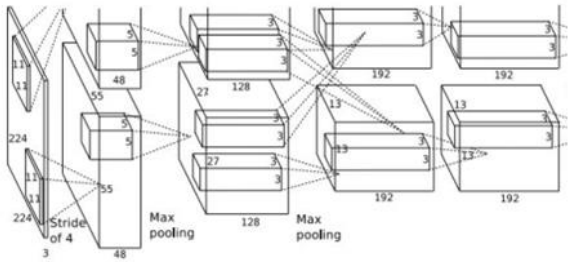
Vaca



Pasto

Altamente ineficiente, repite trabajo ya hecho en detecciones anteriores

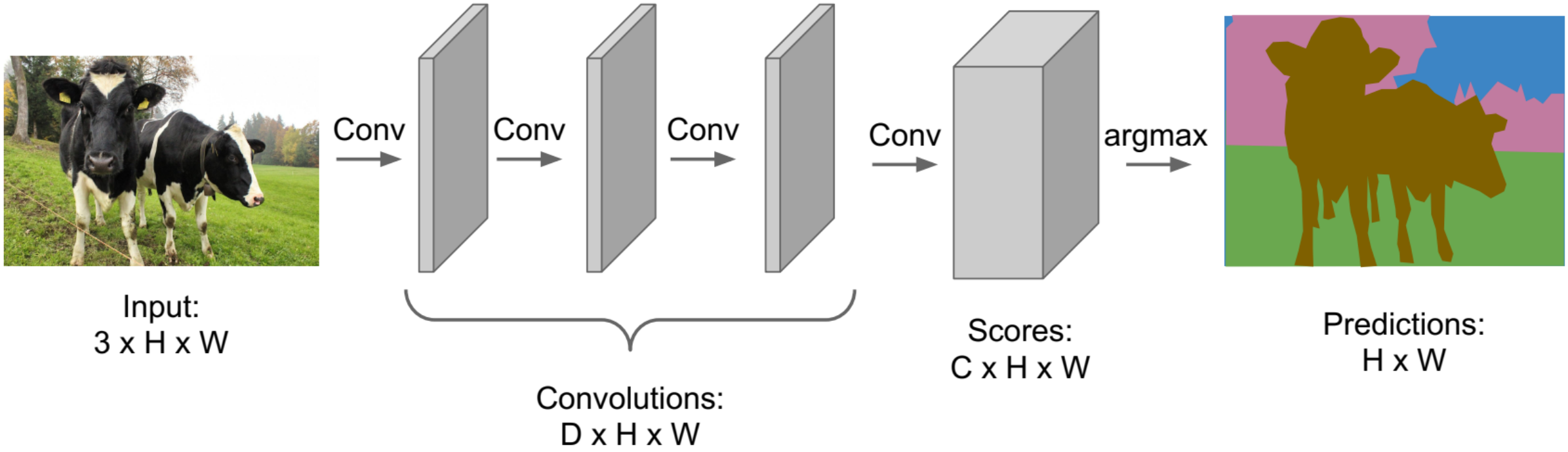
# ¿Cómo usar CNN para segmentación semántica?




Resolución es fuertemente reducida debido al proceso de la red

# ¿Cómo usar CNN para segmentación semántica?

Usamos CNN con filtros pequeños, sin pooling, con padding y manteniendo stride pequeño

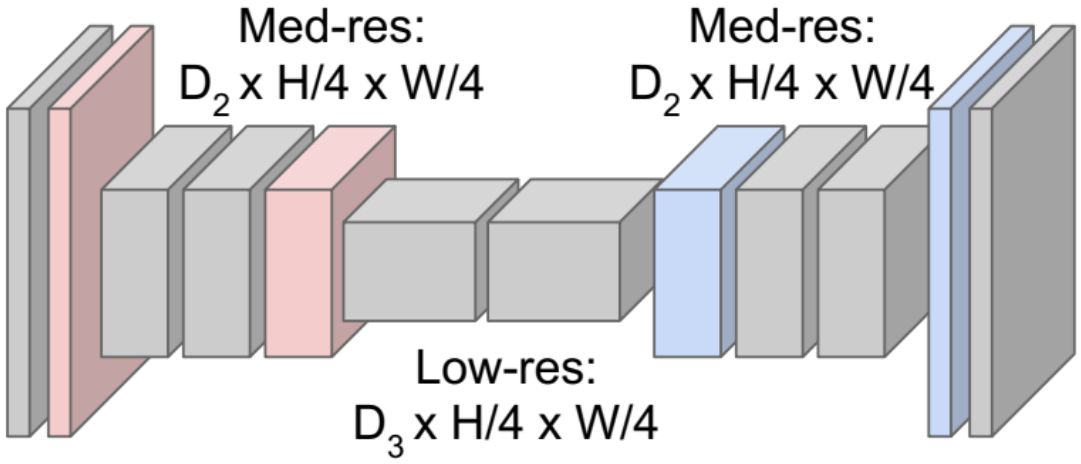



Excesivo costo a nivel de cómputo (no hay pooling)

# ¿Cómo usar CNN para segmentación semántica?

Arquitectura con cuello de botella

Input:
3 x H x W




High-res:
D1 x H/2 x W/2

High-res:
D1 x H/2 x W/2

Predictions:
H x W

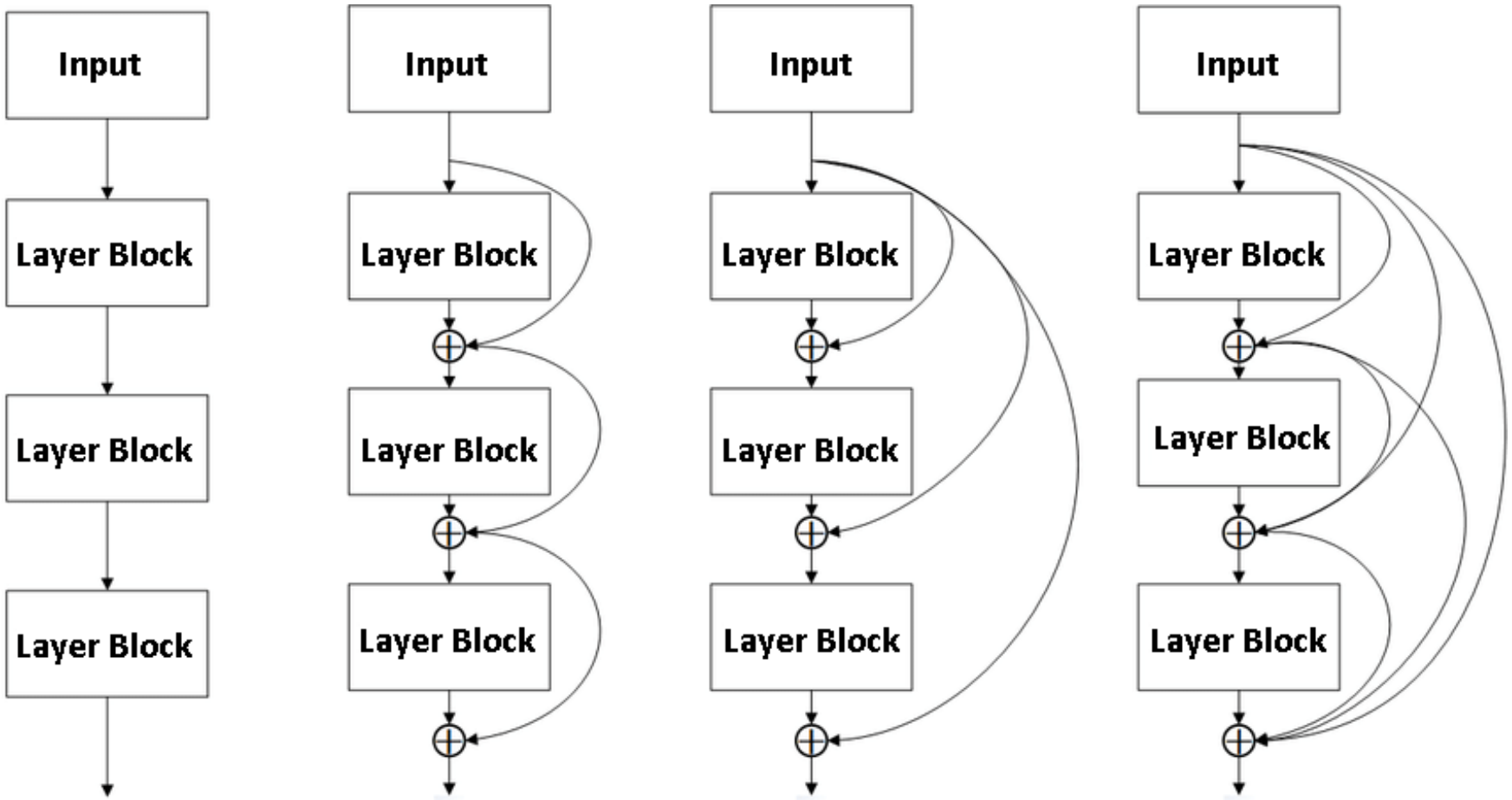
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Clave: convoluciones (downsampling) seguidas de convoluciones transpuestas (upsampling)

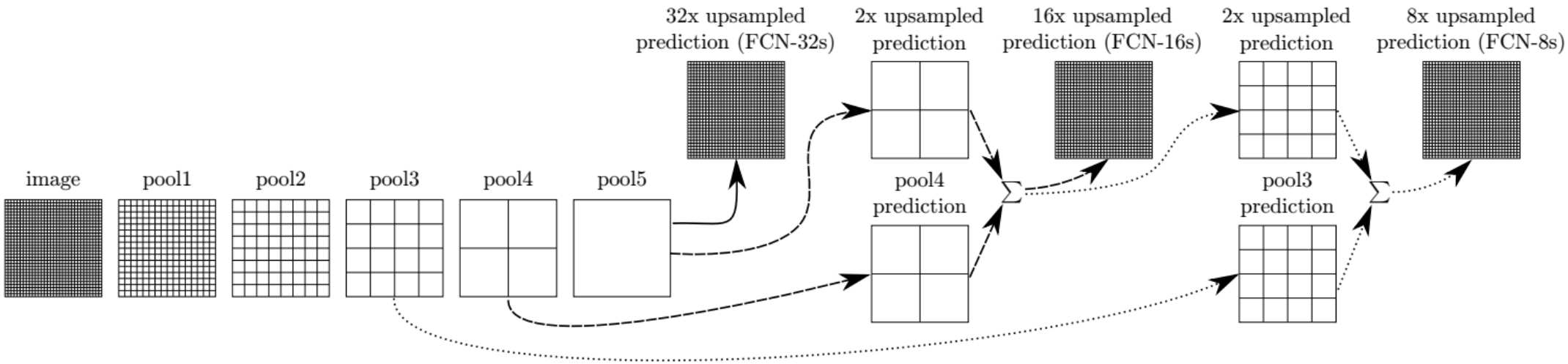
# ¿Y qué tal los resultados?


Bien malos en realidad

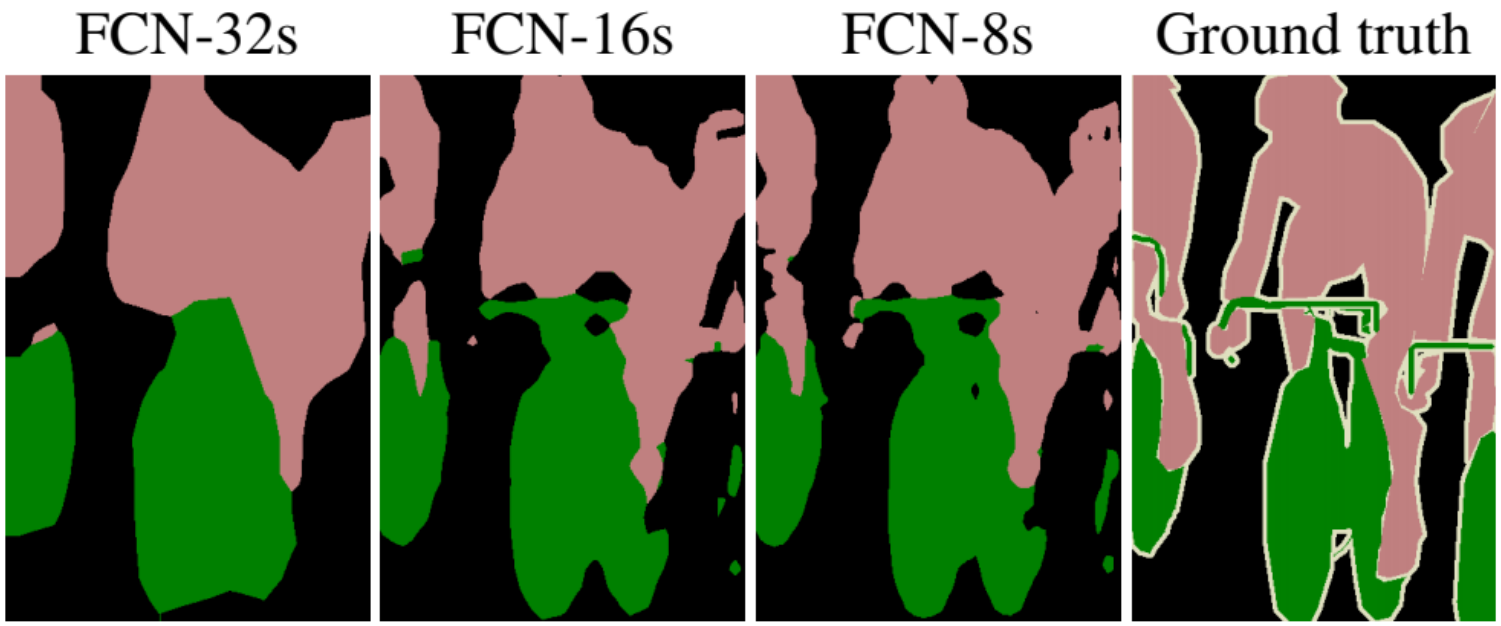
# Interludio: *skip connections*



# Podemos mejorar la resolución utilizando *skip connections*



Podemos mejorar la resolución utilizando *skip connections*



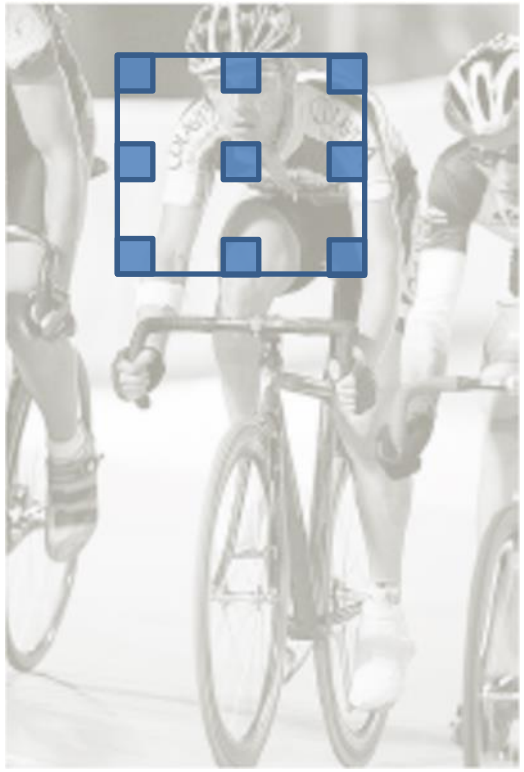
# Actualmente se incorporan más elementos a este pipeline

- Convoluciones dilatadas

# Convoluciones dilatadas entregan mayor contexto espacial

Yu, F., Koltun, V. (2016). Multi-Scale Context Aggregation by Dilated Convolutions. International Conference on Learning Representations (ICLR).

# Convoluciones dilatadas entregan mayor contexto espacial



Yu, F., Koltun, V. (2016). Multi-Scale Context Aggregation by Dilated Convolutions. International Conference on Learning Representations (ICLR).

# Actualmente se incorporan más elementos a este pipeline

- Convoluciones dilatadas
- Mayor profundidad

# U-net refinan idea de *skip-connections* para hacer predicción *coarse-to-fine* usando más capas

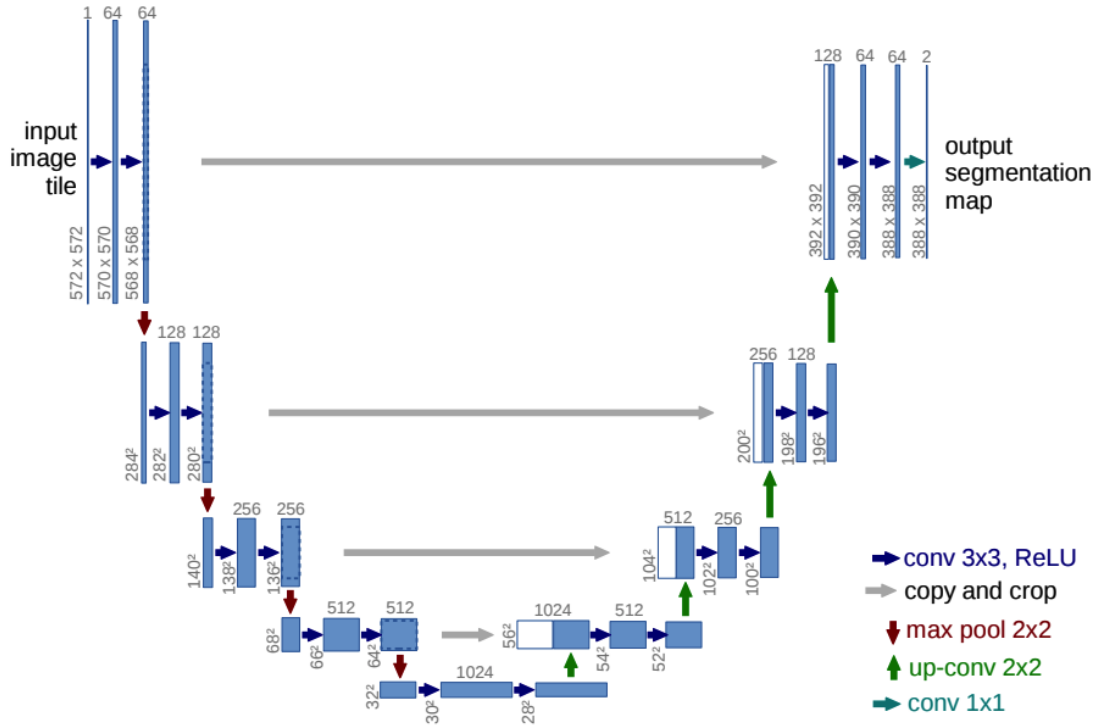


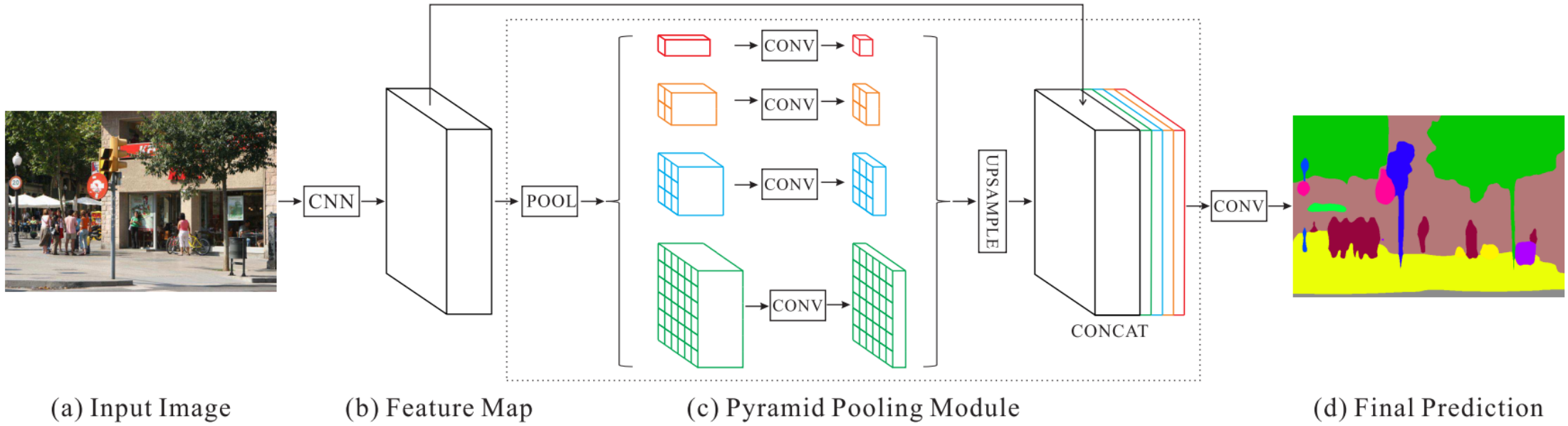

**Fig. 1.** U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Ronneberger, O., Fischer, P., and Brox, T. (2015) "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention.

# Actualmente se incorporan más elementos a este pipeline

- Convoluciones dilatadas
- Mayor profundidad
- Subdivisión estructurada de la imagen

# Subdivisión estructurada de las imágenes permite procesamiento simultáneo en múltiples resoluciones



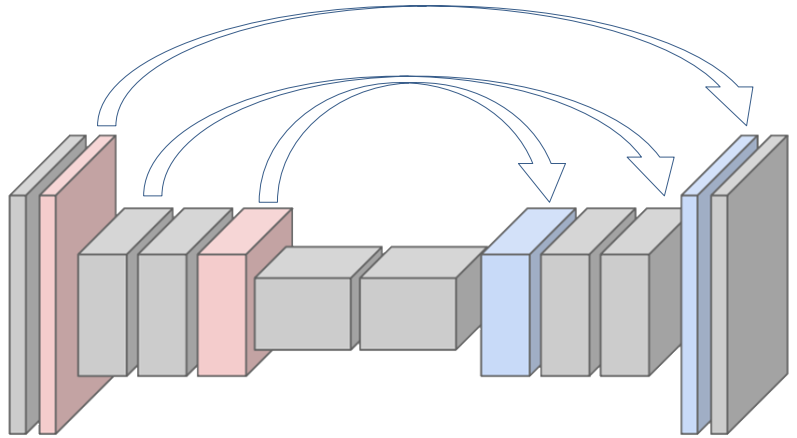

Zha, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Multi-Scale Context Aggregation by Dilated Convolutions. Conference on Computer Vision and Pattern Recognition (CVPR)

https://youtu.be/HYghTzmbv6Q

Zha, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Multi-Scale Context Aggregation by Dilated Convolutions. Conference on Computer Vision and Pattern Recognition (CVPR)

# Resumen de la primera parte

- Es posible utilizar CNNs para analizar entornos urbanos de manera efectiva y útil (p.ej. mapeo de percepción visual).
- Para “entender” lo que ve la red, es necesario plantear la tarea visual como segmentación semántica.
- Para obtener resultados de buena resolución, CNNs deben incorporar numerosos cambios: convoluciones transpuestas, skip-connections, convoluciones dilatadas, etc.
- Resultados recientes muestran gran calidad, segmentando múltiples categorías.



Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería de Transporte y Logística


# Sistemas Urbanos Inteligentes

## Análisis visual de entornos urbanos

Hans Löbel
Dpto. Ingeniería de Transporte y Logística
Dpto. Ciencia de la Computación