



---

# Desarrollo de un método de predicción de puestos laborales mediante redes neuronales convolucionales sobre imágenes satelitales

PAULINA PIZARRO KUMPF

---

SUPERVISORES:

HANS LÖBEL

JUAN CARLOS MUÑOZ

Seminario DITL 2020

---

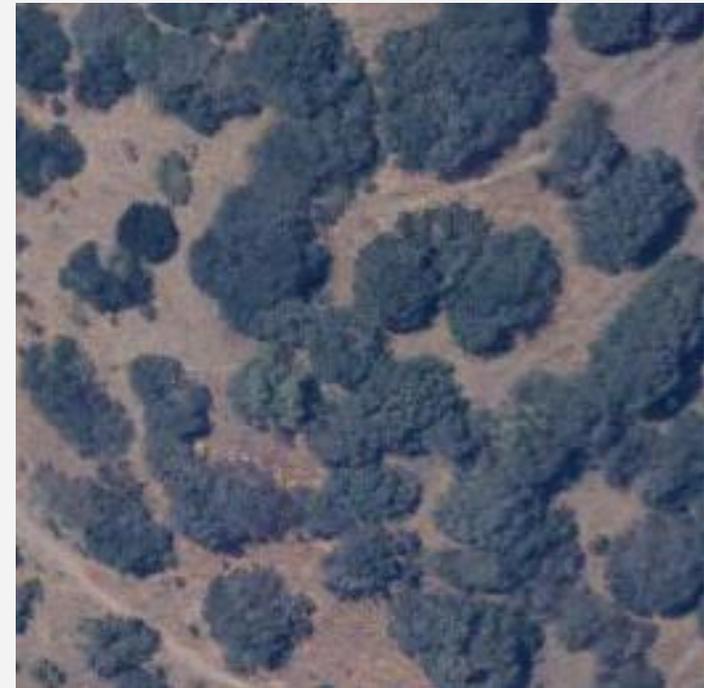
# Participación del público

---



Viendo la imagen cuántas personas creen que trabajan en esta zona?

Porción de una imagen satelital de Santiago



100 m

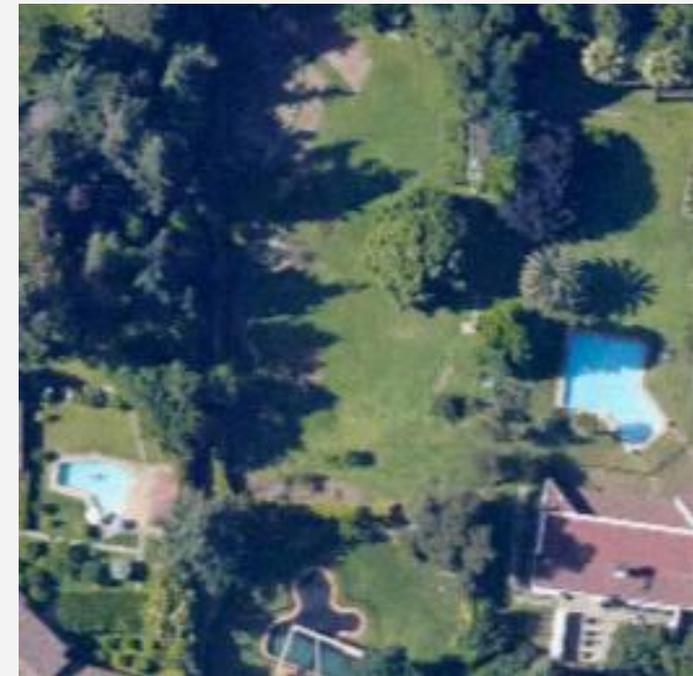
# Participación del público

---



Viendo la imagen cuántas personas creen que trabajan en esta zona?

Porción de una imagen satelital de Santiago



100 m

# Participación del público

---



Viendo la imagen cuántas personas creen que trabajan en esta zona?

Porción de una imagen satelital de Santiago



100 m

# Contenido

---

## Introducción

- Motivación
- Experiencia
- Objetivos

## Metodología

## Experimentos preliminares

## Pasos a seguir



# Introducción - Motivación

---

## Localización de puestos laborales

- Ayudan a comprender desplazamientos dentro de la ciudad
  - Principal propósito de viaje
- Indicador de como crece y se desarrolla la ciudad
  - Economía
  - Forma urbana

Para lograr un [desarrollo urbano sustentable](#), los problemas de sistemas de transporte urbano y la forma urbana de la ciudad son lo primordial (Kenworthy, 2006)

# Introducción - Motivación

---

Fuentes de información de localización de puestos laborales en Santiago:

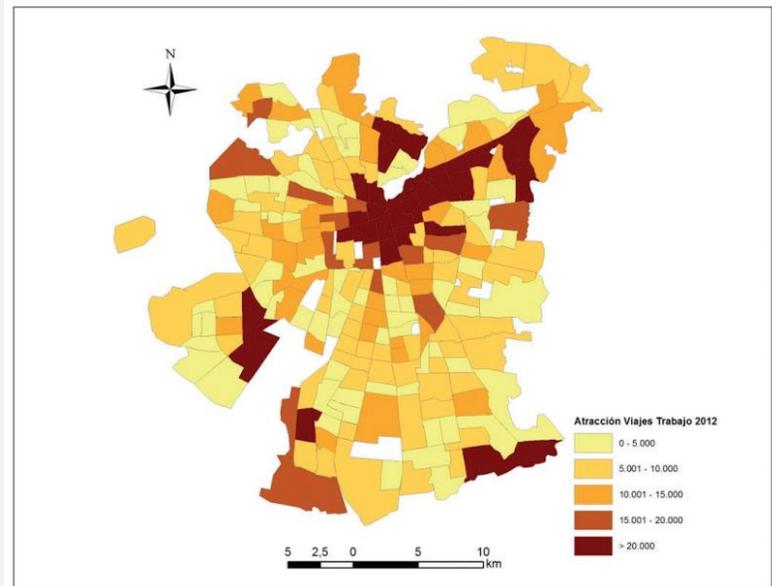
- Encuesta Origen Destino: Coordenadas de destinos con motivo laboral declarado por encuestados
- Servicio de Impuestos Internos: Cantidad de trabajadores por empresa asignados a su casa matriz

Fuente	EOD	SII
Alcance	Un bajo porcentaje de los trabajos, incluyendo informales (~18.000)	El total de los trabajos formales (~4 millones en 2012)
Última actualización	2012	2019 (Se obtiene todos los años)
Precisión de las localizaciones	Alta	Media (Asignaciones solo a casa matriz)

# Introducción - Motivación

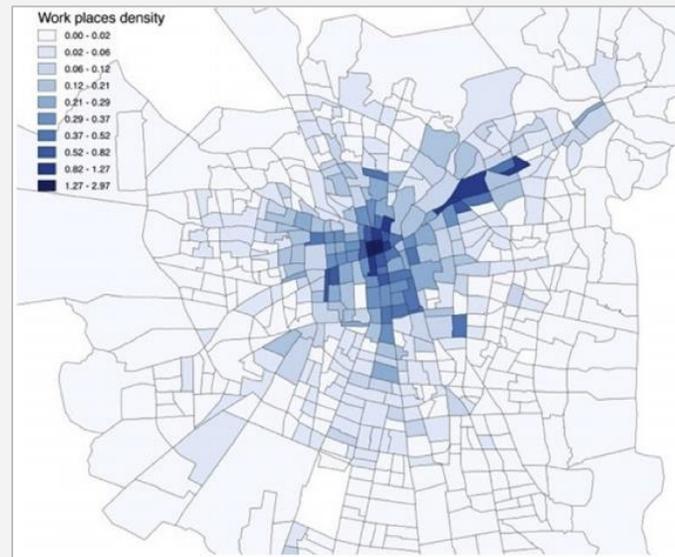
Cómo se usa y procesa la información puestos laborales actualmente?

En base a Encuesta Origen destino:



(Contreras, Navarrete y Arias, 2017)

En base a uso de suelos SII:



(Niehaus, Galilea y Hurtubia, 2016)

Actualmente se obtienen **indicadores** de cantidad de puestos de trabajo por zonas geográficas.

Los indicadores sirven como forma de comparación entre zonas pero no individualmente como número de trabajos existentes en una zona.

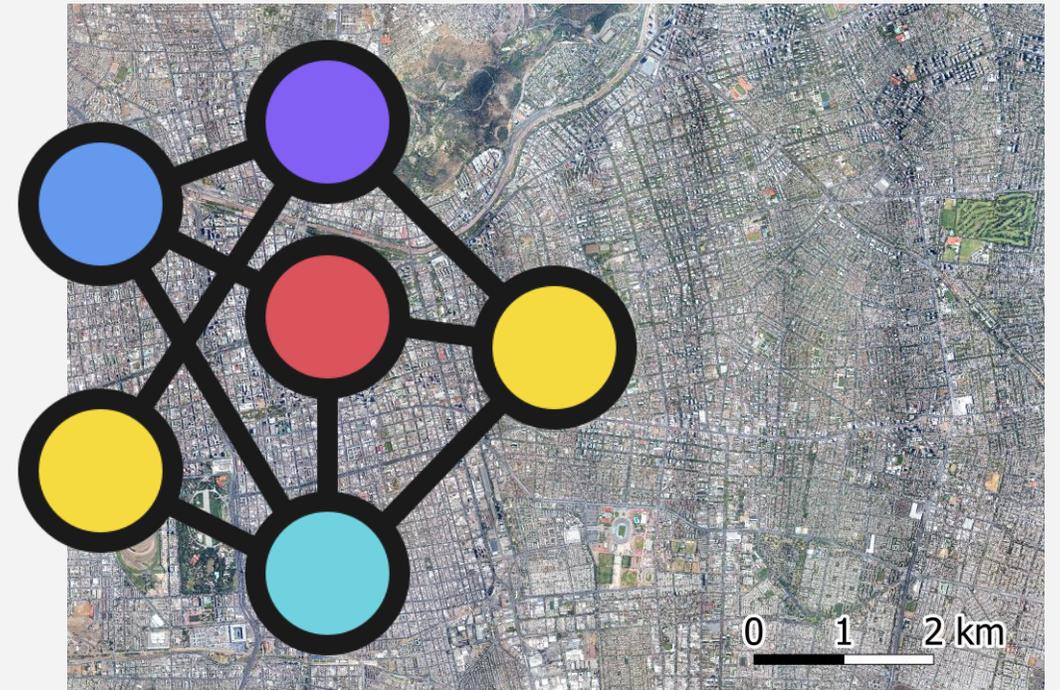
# Introducción - Motivación

---

Existe una forma de obtener y actualizar la localización de puestos laborales de forma más detallada a un costo razonable?

Imágenes satelitales: Fuente de datos sobre la ciudad actualizada periódicamente

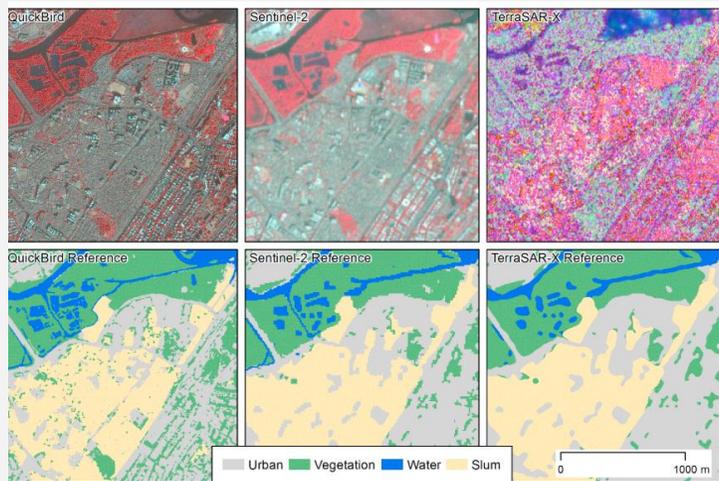
*Machine Learning*: Comúnmente usado para analizar y procesar grandes cantidades de datos.



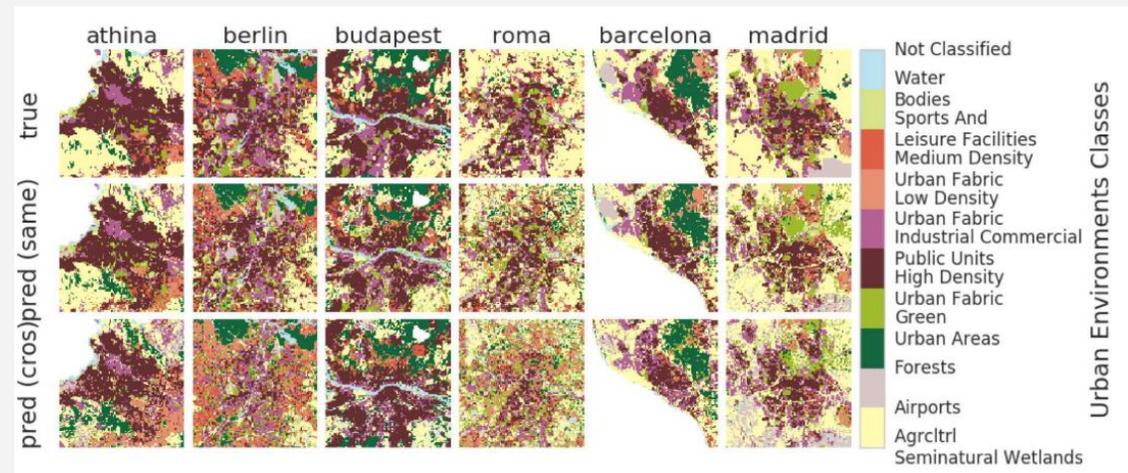
# Introducción - Experiencia

## Uso de *machine learning* sobre imágenes satelitales

- Identificación de tipo y uso de suelos: Agua, vegetación, urbano y slums (urbano con pobreza altamente denso) (Wurm et al , 2019).
- Identificación patrones de entornos urbanos (Albert, Kaur y Gonzalez, 2017)



(Wurm et al , 2019)

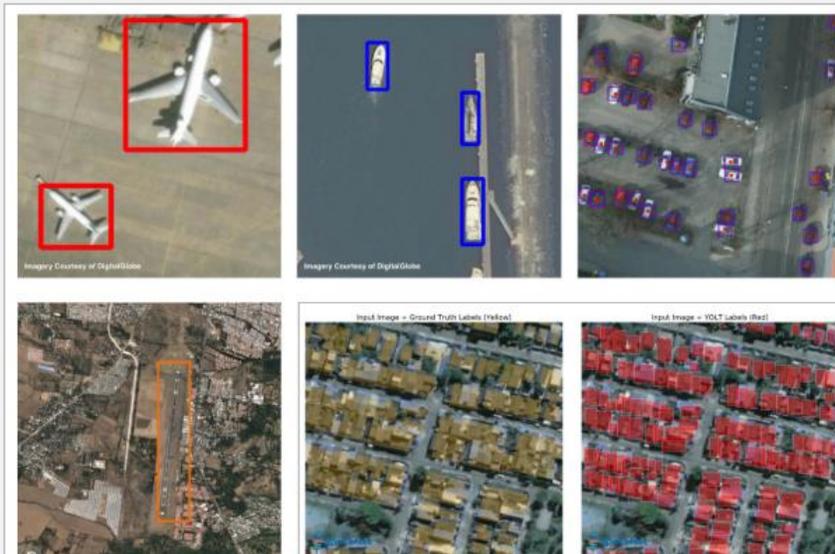


(Albert, Kaur y Gonzalez, 2017)

# Introducción - Experiencia

## Uso de *machine learning* sobre imágenes satelitales

- Detección de autos, aviones y barcos (Van Etten, 2018)
- Extracción de carreteras y caminos para ayuda humanitaria (Bonafilia et al., 2019)
- Mapeo de asentamientos de refugiados (Quinn et al., 2018)



(Van Etten, 2018)



(Bonafilia et al., 2019)

# Introducción - Objetivos

---

Desarrollo de un método de predicción de localización de **puestos laborales** utilizando **redes neuronales** convolucionales sobre **imágenes satelitales**

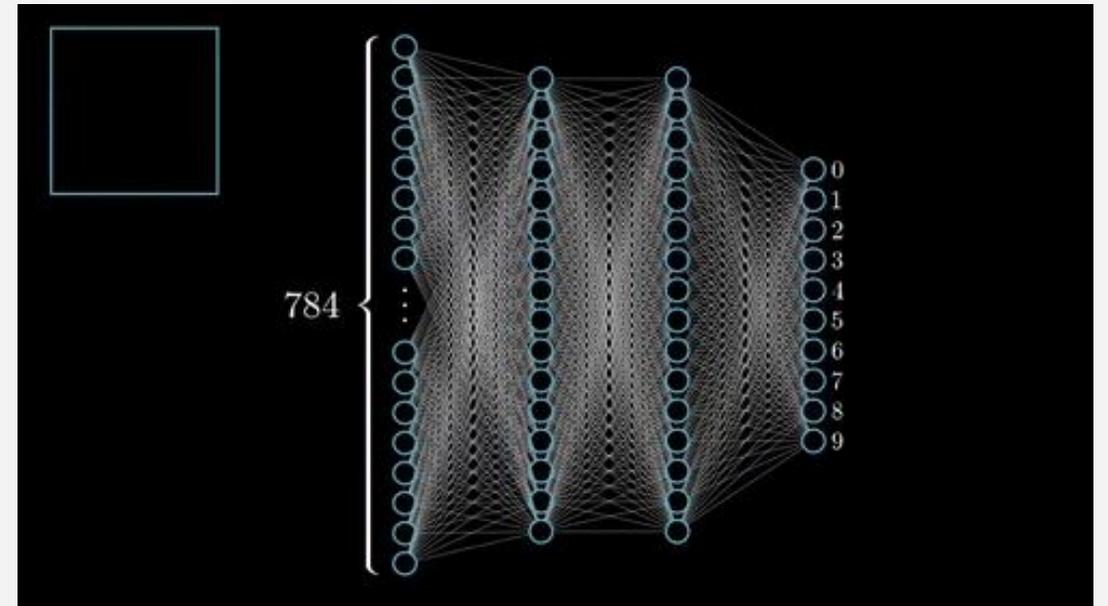
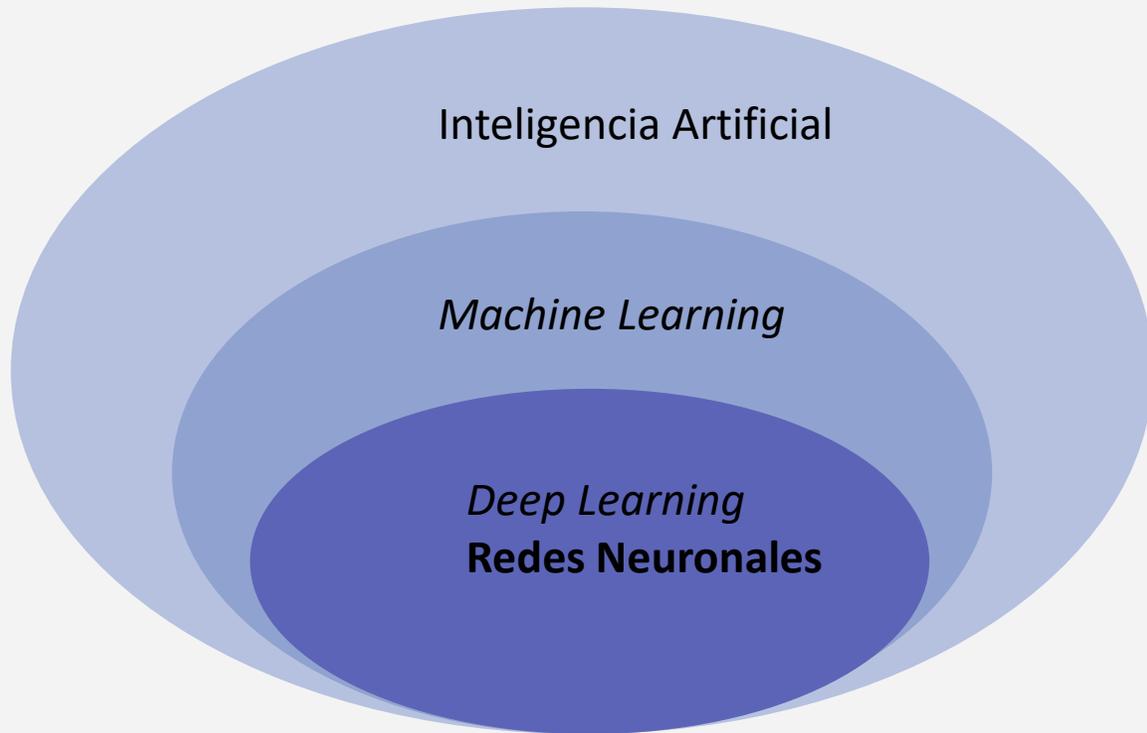
## Contribuciones

- Cumplir con ventajas comparativas sobre las metodologías actuales:
  - Menos costoso de aplicar
  - Detalle
- Aportar en interdisciplina entre ciencias de la computación e ingeniería de transporte

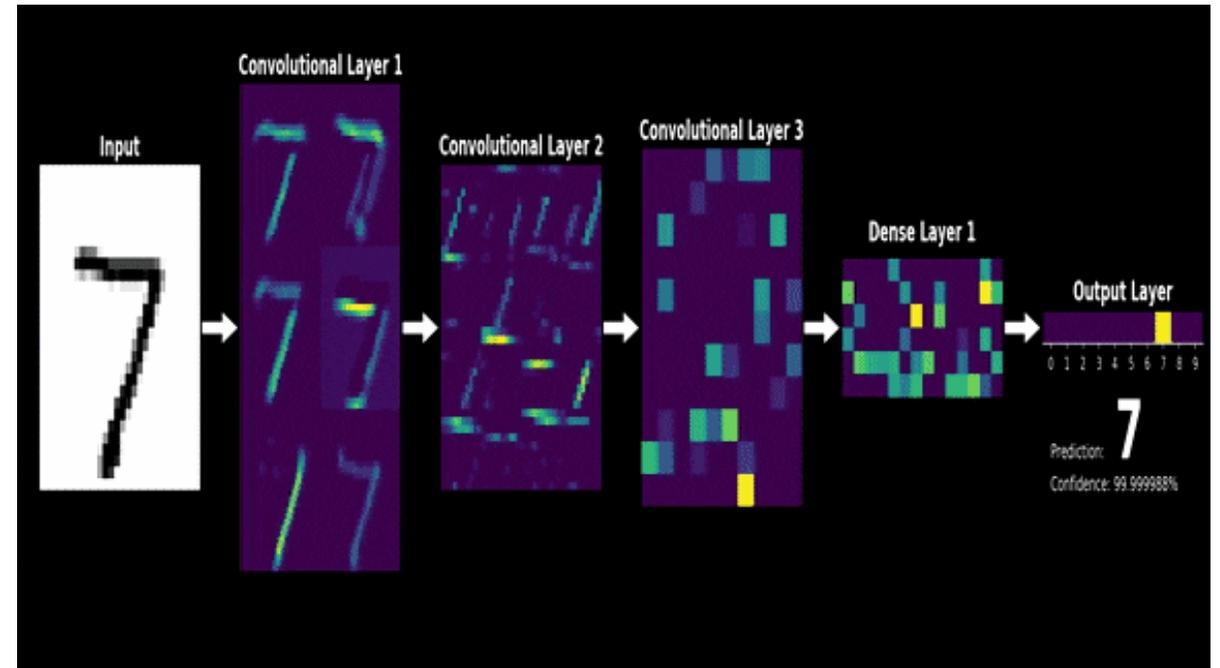
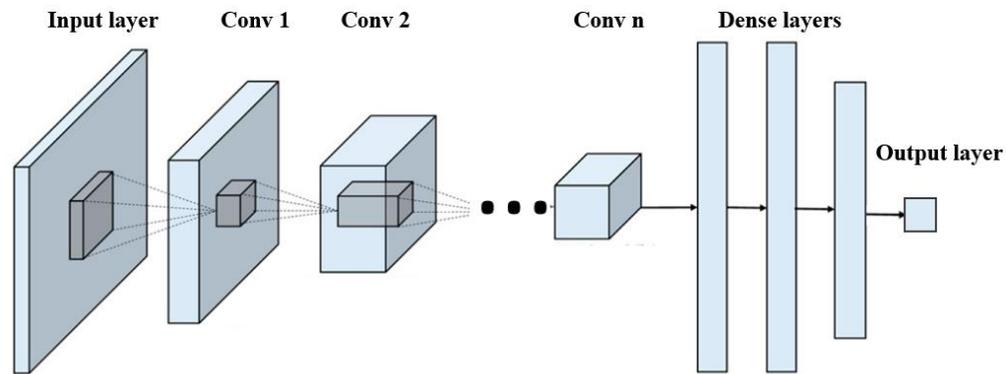
# Metodología



# Metodología – Redes Neuronales



# Metodología – Redes Neuronales Convolucionales



# Metodología – Inputs y outputs

---

## Imagen satelital de alta precisión de Santiago



(Catálogo Nacional de Información Geoespacial, 2015)

- 1 pixel equivale a 0,36 metros
- Archivo pesa 70 GB

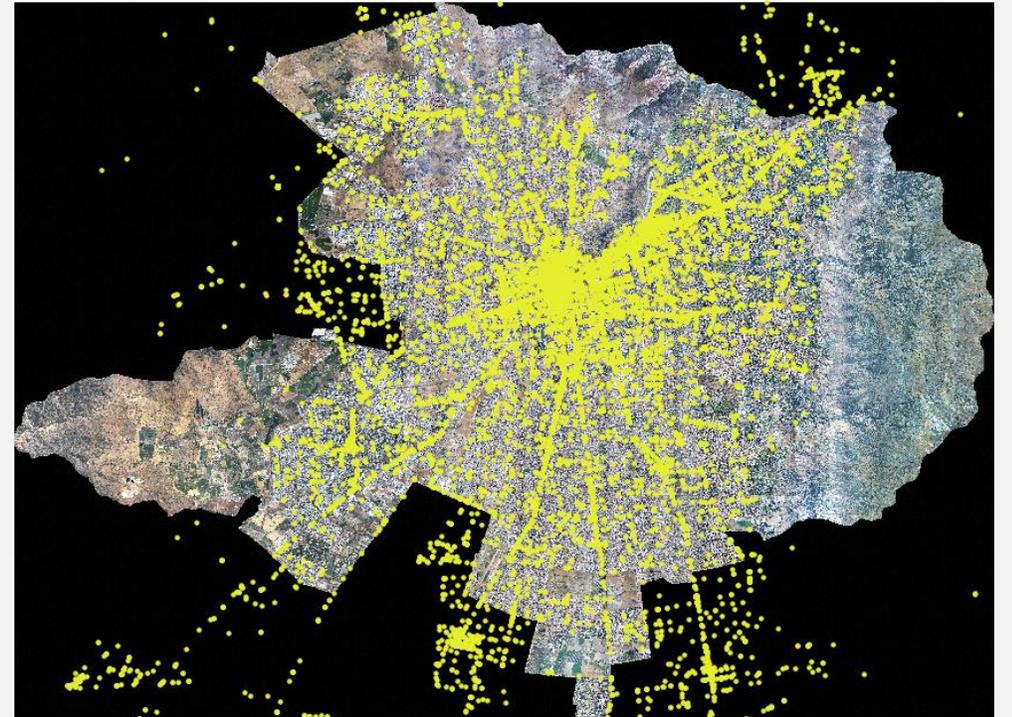
# Metodología – Inputs y outputs

---

## Localización de trabajos según EOD 2012

Trabajos obtenidos de georreferenciación de destinos de viajes con propósito al trabajo

Contabilizando los correctamente referenciados se obtuvieron **17.256** trabajos



# Metodología – Inputs y outputs

---

## Localización de trabajos según SII

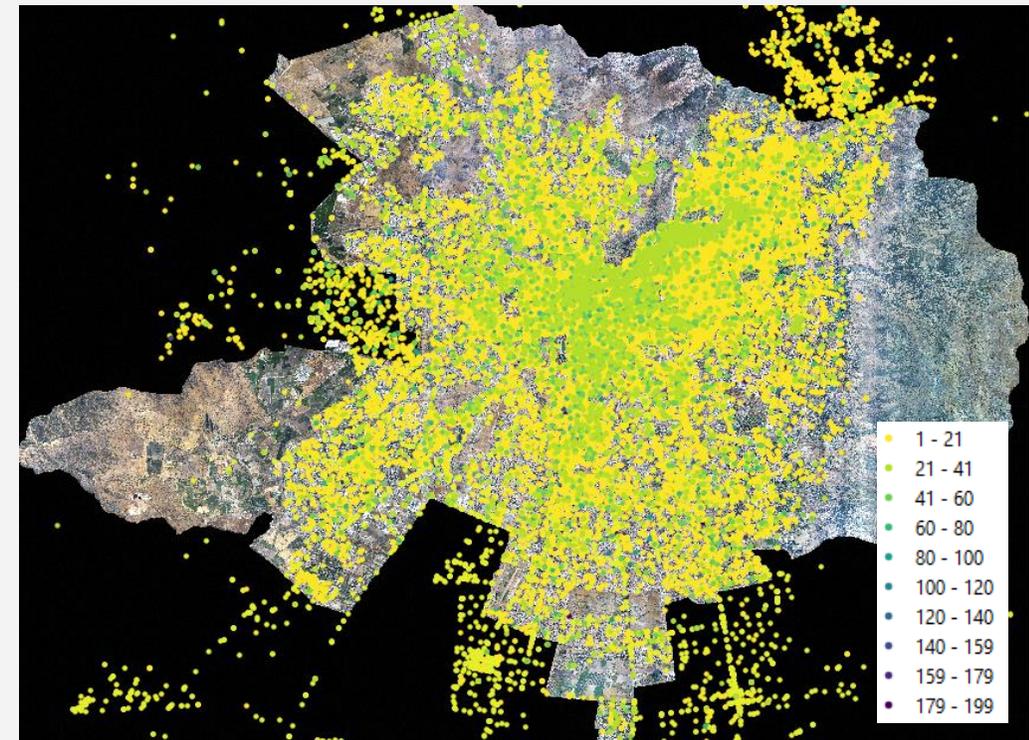
- Registro de empresas con dirección de su casa matriz y cantidad de empleados
- Un total de 4.408.208 de trabajadores a empresas de Santiago en 2012
- Solo se cuenta con ubicación de casa matriz y no otras sedes donde trabajadores ejercen funciones si es que tienen
  - Empresas con mayor cantidad de trabajadores son más posibles de tener otras sedes

Se tomaron en cuenta solo empresas con menos de 200 trabajadores obteniendo un total de **1.032.278** trabajos georreferenciados

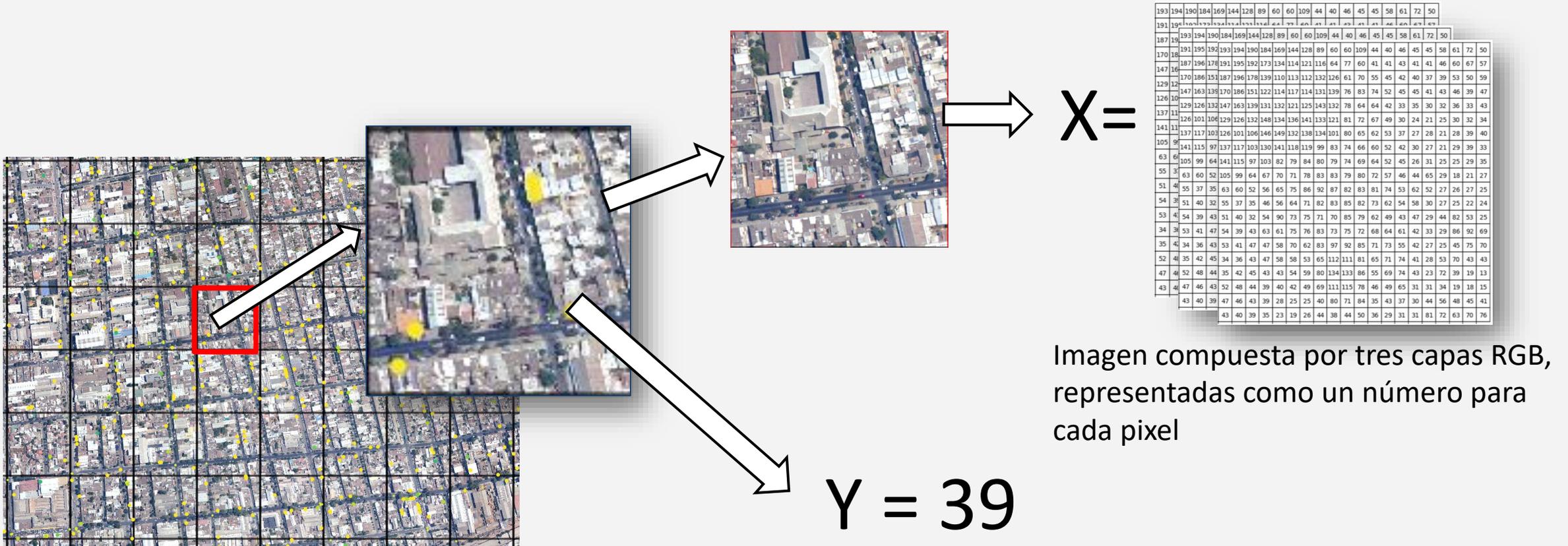
# Metodología – Inputs y outputs

## Localización de trabajos de EOD y SII

	Cantidad de trabajos obtenidos	Descripción
EOD	17.256	Viajes con propósito al trabajo correctamente georreferenciados
SII	1.032.278	Empresas con menos de 200 trabajadores con direcciones georreferenciables. De un total de 4.408.208
Total	1.049.534	



# Metodología – Inputs y outputs



# Metodología – Inputs y outputs

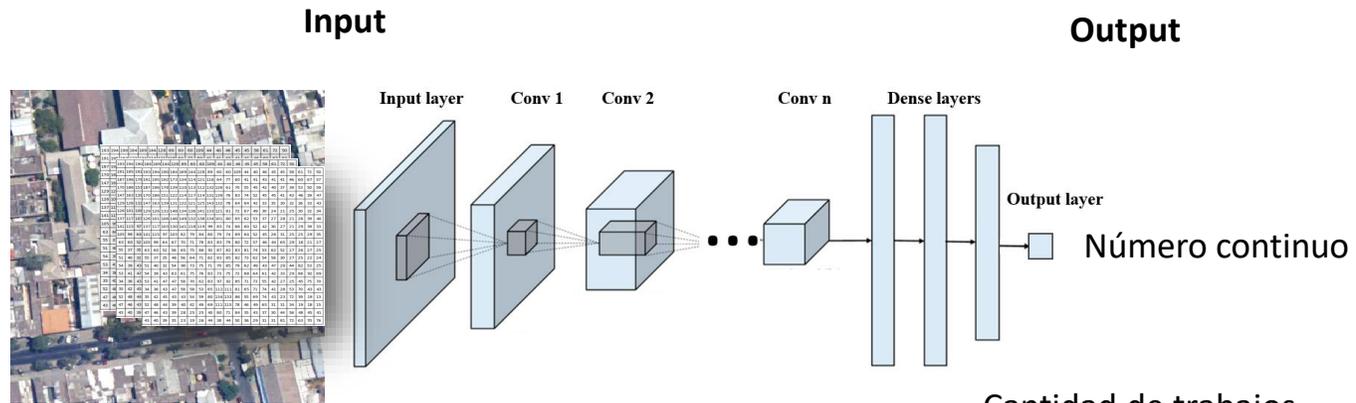


Imagen representada por tres set de números asociados a cada pixel en su respectivo color RGB

Cantidad de trabajos asociados a cada imagen

Función de pérdida / F.O.

Entrenamiento minimiza el error cuadrático medio entre la predicción y el dato entregado.

Cantidad de parámetros

Con cada capa\* se añaden parámetros al modelo. La red con la que trabajamos ahora tiene 271.713

# Metodología – *Overfitting* y cómo evitarlo

*Overfitting* (sobre entrenamiento) ocurre cuando un modelo predice con alta precisión un set particular de datos pero falla en predecir ante nuevos datos.

En el contexto de esta investigación utilizamos tres mecanismos para evitarlo:

- Gran cantidad de datos:

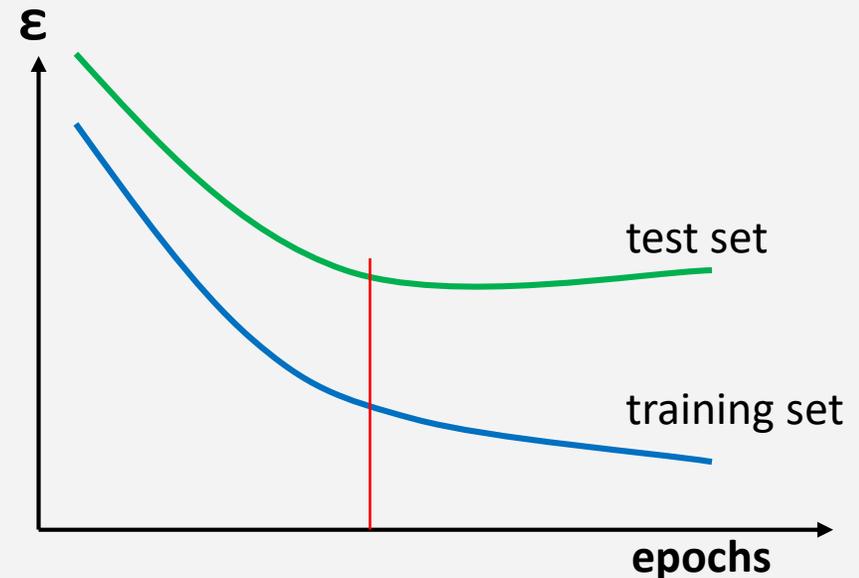
La segmentación en cuadrados de la imagen satelital permite obtener más de 100.000 imágenes para entregar al modelo.

- Aumentación de datos:

Rotación aleatoria de imágenes para aumentar variedad.

- Uso de *dropout*:

*Dropout*, aplicado a una capa, de forma aleatoria reemplaza atributos con 0.

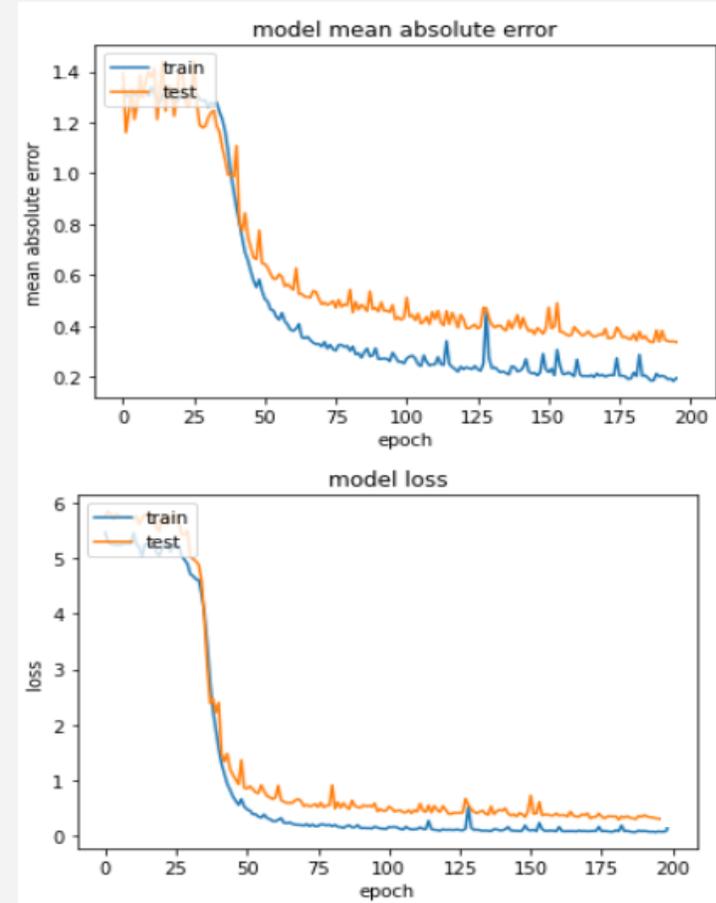


Esperimento preliminar



# Experimento preliminar

- 15.000 imágenes
- Entrenamiento sobre trabajos de la EOD
- Función de pérdida de error cuadrático medio
- Uso de servicio Google Colab
- Tiempo de cómputo: 4 horas



# Experimento preliminar - Resultados

---

Error absoluto medio sobre set de entrenamiento: 0,17

Error absoluto medio sobre set de prueba: 0,29

## Mejores predicciones sobre set de prueba



Elaboración propia

- Identifica terrenos no construidos y les asigna 0 trabajos
- Asigna trabajos correctamente a áreas urbanas con pocos trabajos

# Experimento preliminar - Resultados

Error absoluto medio sobre set de entrenamiento: 0,17

Error absoluto medio sobre set de prueba: 0,29

- A mayor cantidad de trabajos fallan más las predicciones
- El set de datos preliminar cuenta con muy pocos trabajos al ser solo los de la EOD, por lo tanto, hay mucha variabilidad y ocurre que imágenes similares tienen cantidades de trabajos muy distintas

## Peores predicciones sobre set de prueba



Elaboración propia

Pasos a seguir



# Pasos a seguir

---

## Lecciones aprendidas sobre experimento preliminar

- Balancear cantidad de imágenes con 0 trabajos
- Entrenar con más imágenes
- Evitar alta correlación entre el set de entrenamiento y set de prueba

## Próximos experimentos

- Uso de cluster IALAB. Mayor capacidad y tiempo de cómputo
- Modelo más complejo
- Set de datos sobre 40.000 imágenes

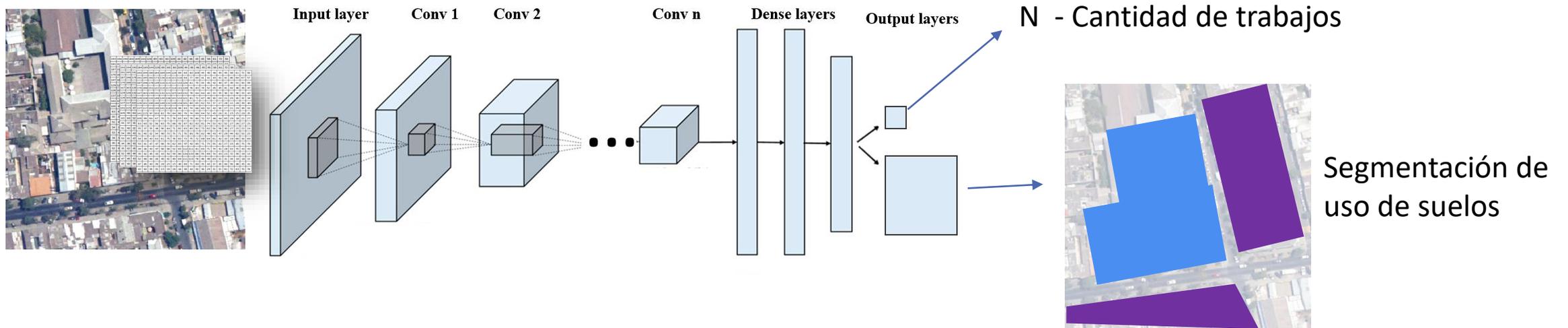


Imágenes comparten grandes áreas

# Pasos a seguir – Red con dos *outputs*

Cantidad de trabajos y uso de suelos son variables altamente relacionadas.

**Hipótesis:** Entrenar la red para predecir ambas simultáneamente puede resultar en una mejor predicción de cantidad de trabajos que modelo con un solo output.



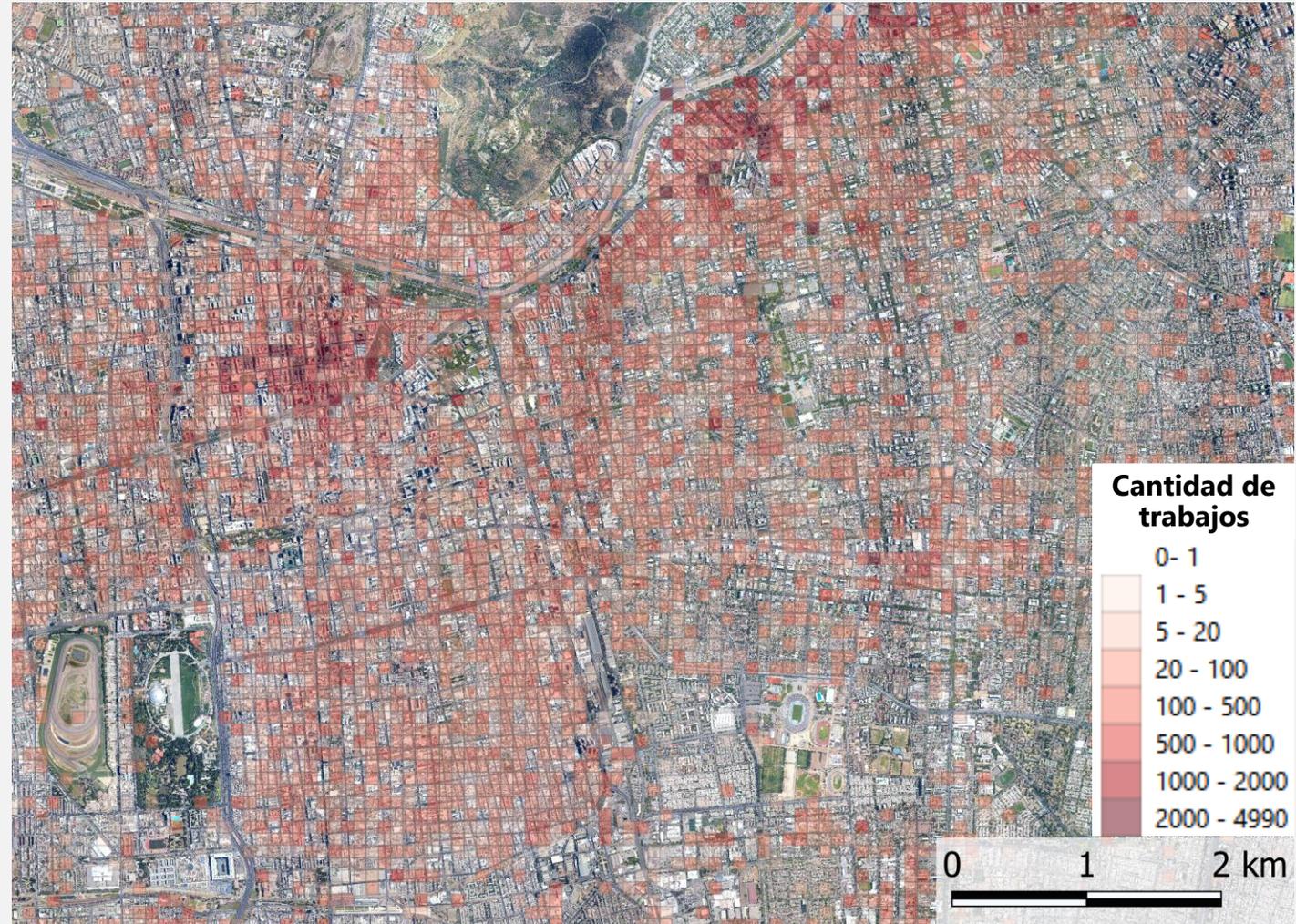
# Pasos a seguir

---

## Resultados esperados

Red neuronal entrenada para usar sobre imágenes satelitales:

- Actualizar cantidad de trabajos por segmentos de Santiago
- Obtener cantidad de trabajos en otras ciudades similares



Elaboración propia sobre imagen de Catálogo Nacional de Información Geoespacial (2015)



---

# Desarrollo de un método de predicción de puestos laborales mediante redes neuronales convolucionales sobre imágenes satelitales

PAULINA PIZARRO KUMPF

---

SUPERVISORES:

HANS LÖBEL

JUAN CARLOS MUÑOZ

Seminario DITL 2020

---

# Bibliografía

---

Kenworthy, J. R. (2006) 'The eco-city: Ten key transport and planning dimensions for sustainable city development', *Environment and Urbanization*, 18(1), pp. 67–85.

Contreras, R., Navarrete, C., & Arias, G. (2017). Encuesta Origen-Destino Santiago 2012 Evolución de escenarios de accesibilidad. *Ingeniería de Transporte*, 20(1).

Niehaus, M., Galilea, P. and Hurtubia, R. (2016) 'Accessibility and equity: An approach for wider transport project assessment in Chile', *Research in Transportation Economics*. Elsevier Ltd, 59, pp. 412–422.

Wurm, M. *et al.* (2019) 'Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks', *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier, 150(May 2018), pp. 59–69.

Albert, A., Kaur, J. and Gonzalez, M. C. (2017) 'Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale', *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Part F1296, pp. 1357–1366.

Van Etten, A. (2018) 'You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery'. Available at: <http://arxiv.org/abs/1805.09512>.

Bonafilia, D. *et al.* (2019) 'Building High Resolution Maps for Humanitarian Aid and Development with Weakly-and Semi-Supervised Learning', *The IEEE Conference on Computer Vision and Pattern Recognition - Workshops*Quinn, J. A. *et al.* (2018) 'Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128).

Ministerio de Economía (2015) *Fotografía aérea del Gran Santiago, Catálogo Nacional de Información Geoespacial*. Available at: <http://www.ide.cl/descargas/capas/economia/Fotografia-aerea-Gran-Santiago.rar>.