# 2018 Data Science Bowl

Cell Nucleus Segmentation for Medical
Research Advancement

Rachit Kataria, Lucas Hu, and Sean Syed

# Why Nuclei?

# Why Nuclei?

- 40% of all deaths are caused by heart disease and cancer

- The average time-to-market for a new drug is 10 years

- Automating this process allows for:
    - Efficient measuring of cells in response to different stimuli
    - Expedited drug research process
    - Reduced time-to-market for new drugs
    - Improved throughput for research and insight

# Machine Learning for Image Segmentation: A Brief Primer

# U-NET

- Convolutional network tailored for biomedical image segmentation
  - A *contracting* path to capture context
  - A *symmetric* expanding path that enables precise localization
- Skip connections combine information from earlier layers with low resolution information of deeper layers
- Speedy!
- Input size agnostic
- Very large number of layers, slow to train
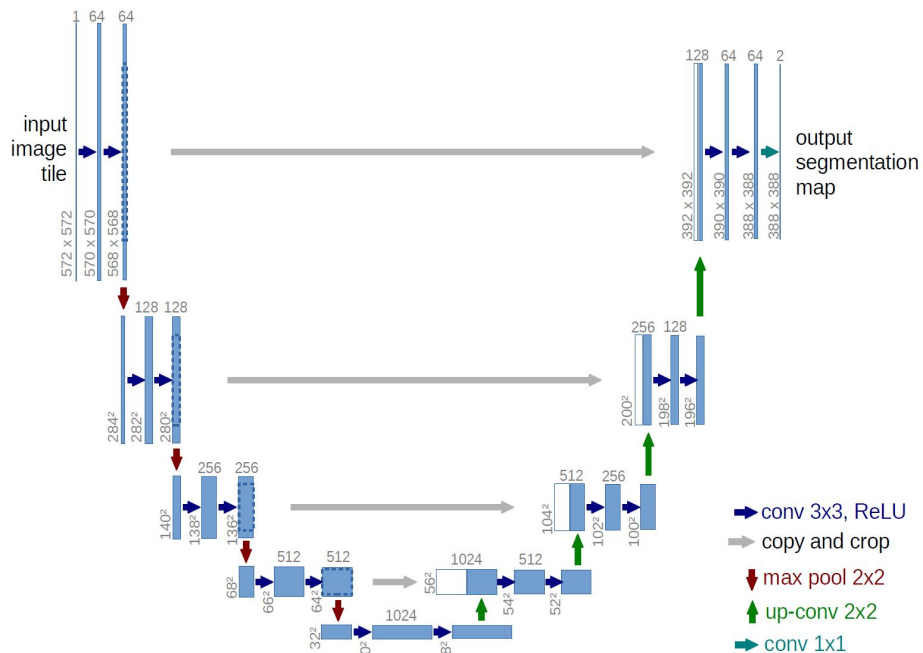
# Mask R-CNN



- Expanding on Faster R-CNN for **pixel** level segmentation
- **Input**: CNN Feature Map
- **Output**: Matrix with 1s on all locations where the pixel belongs to the object and 0s elsewhere
- RoI (Region of Interest) Pool → RoIAlign
    - Avoid rounding down, instead use <u>bilinear interpolation</u>

Competition Approach

# Mask R-CNN vs U-Net

- Divided and conquered (sort of)

- U-Net is easier to implement and understand, but Mask R-CNN was the most effective approach

- Small tweaks made huge differences

# Results &
# Lessons Learned

# 0.468

Mean taken over the individual average precisions at
IoU of each image in the test dataset

# Top 10%

CAIS++ making moves for the second year in a row!

# Key Takeaways

1. Kaggle is a fantastic way to apply machine learning to **real-world problems**

2. A lot of ML is just **tweaking other people's models**

3. **Understanding the code** is half the battle