



# DATA SCIENCE AND MACHINE LEARNING SPECIAL INTEREST GROUP

February 2022 Meeting

# Who we are...

## Mission Statement:

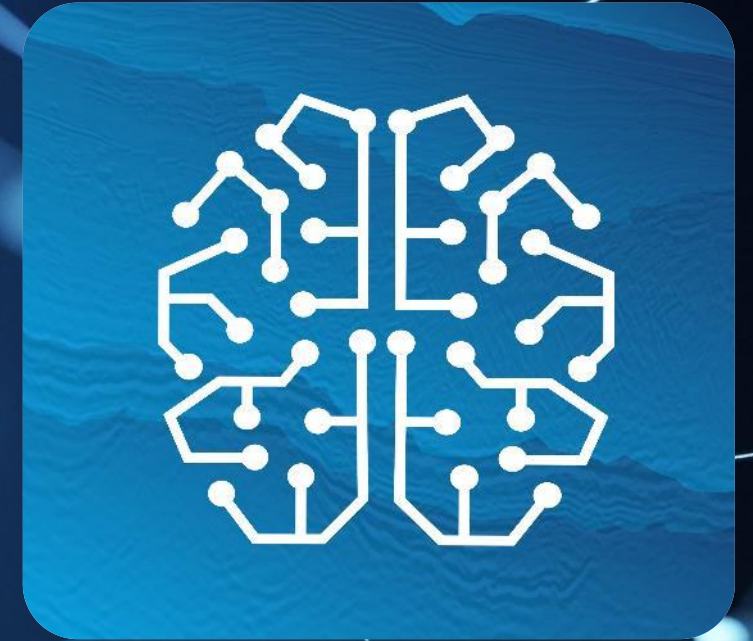
“To build an ecosystem that fosters the development of data science and machine learning, as tools to drive innovation in geosciences.”

## What do we want to accomplish?

- Facilitate industry adoption of tools and workflow that leverage data science and machine learning.
- Help develop strong domain knowledge and data science skills in the geoscience community
- Encourage collaboration between data scientists and geoscientists

## How do want to do that?

- Monthly technical talks from industry professionals
- 1 -2 competitions per year to strengthen key skills
- Annual symposiums to showcase student research
- A digital platform for collaboration and networking (SLACK)





# SIG Committee Members



Eduardo Alvarez  
Software Engineering  
Manager  
Quantico Energy Solutions

Committee Chair



T. Altay Sansal  
Data Science, Lead  
TGS

Special Events



Ben Lasscock  
Technical Lead  
TGS

Technical Talks



John O'Donnell  
Data Scientist  
UH HPE Data Science Institute

Social Events



Ayodeji Babalola  
Hewlett Packard Enterprise

Community Support



We collaborate with clients in their digital transformation initiatives to create possibilities that deliver orders of magnitude performance improvements – in business results and expert efficiency.



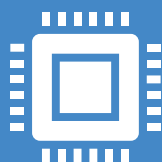
Materials Science and  
Chemistry



Energy



Life Sciences



Semiconductors



Powerful subsurface modeling solutions that deliver on today's most pressing issues in the oilfield: real-time, higher resolution, sustainability, and scalability.



Machine Learning  
Solutions



Drilling Optimization



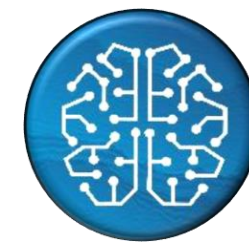
Earth Modeling



Synthetic Well Logging



# Geo2Data Workshop



Salt Identification Challenge - CNNs  
11:00 – 12:00am CT  
February 9<sup>th</sup>, 2021

Our Geo2Data Workshops will focus on building up your practical skills and developing a strong data science foundation. Our goal with these workshops, is not to help you create the next BIG THING, it is to give you practical exposure. Additionally, participants will join a network of aspiring data scientist in experiencing one of the discipline's key rites of passage, hackathons.

**Mentors:** John O'Donnell, Altay Sansal, and Eduardo Alvarez



## Geo2Data Workshop: Salt Identification Challenge – CNNs

In this workshop we look back to some of the more comprehensive notebooks from the **2019 TGS Salt Identification Challenge**. This challenge took place 3 years ago and was the first of its kind for our discipline.

***11:00pm – 12:00pm CT***



**#SlackHack** ***12:00pm – 12:30pm CT***

THANK YOU TO OUR SPONSORS

QUANTICO  
ENERGY SOLUTIONS



# The conversation continues on SLACK . . .

Exclusive invitations will be sent to all attendees of our webinars to join our SIG Slack group!



Digital Networking



Engage Thought Leaders



SIG Announcements



Talent Discovery

Much More...



[data-science-gsh.slack.com](https://data-science-gsh.slack.com)



# Event Procedure



Please stay muted throughout the event



No recording or screen captures of the speaker's content



Questions will be addressed at the end of the presentation by the speaker and moderator



Trivia questions may appear in transition points throughout the presentation. You are not obligated to participate, and all answers are anonymous.



At the beginning of the Q&A section, you will have the opportunity to give the speaker and the committee feedback about the event through a simple survey. We highly encourage you to participate so that we can continue to improve our events.



# Geo2Data

## Salt Identification Challenge - CNNs

---

Our goal with these workshops, is not to help you create the next BIG THING, it is to give you practical exposure.



# Workshop Outline



Review of High-Level Concepts



#SlackHack Introduction



Salt Identification – Notebook



Q&A



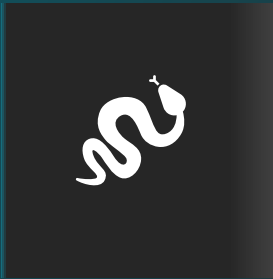
#SlackHack Kick-Off

## What we hope to achieve?

- Getting ML Ready
- Salt Identification
- Practical Understanding of CNNs
- What are U-Nets?
- Data Augmentation
- Apply learnings to real world data

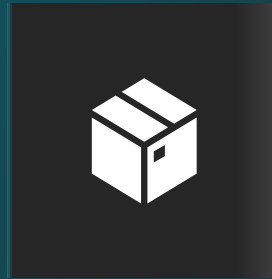
# Learning Objectives

# Getting ML Ready



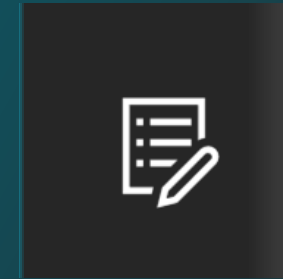
## Anaconda and Miniconda

Python developers go to package, dependency, and environment management system.



## Dedicated Environments

By using them, it is possible to create multiple environments, each one with different versions of packages.



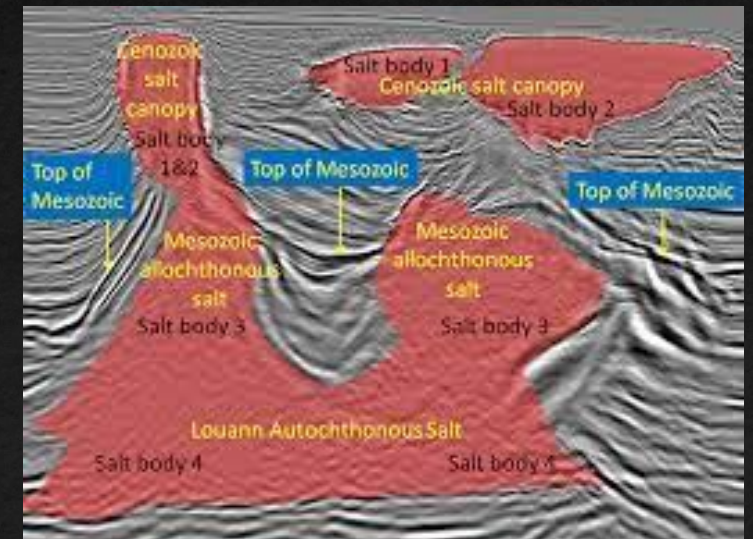
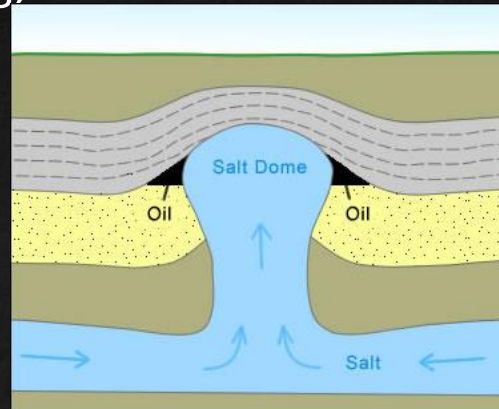
## Working in Notebooks

A simple and elegant editor for experimentation, testing, and prototyping.

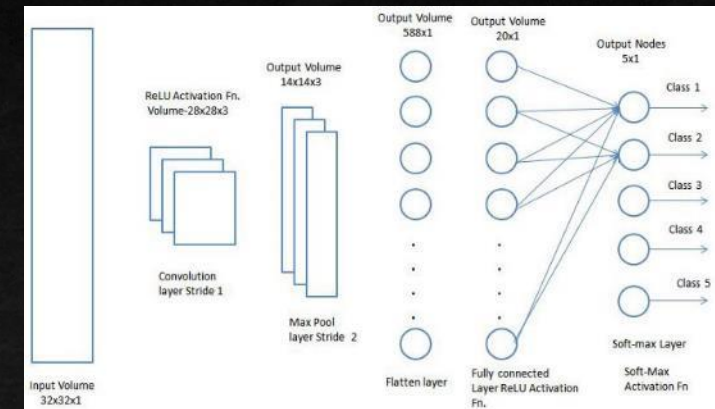
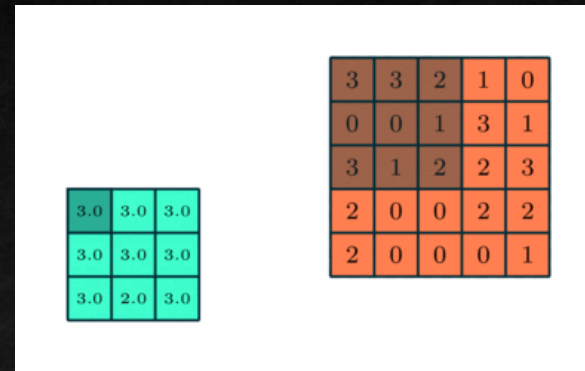
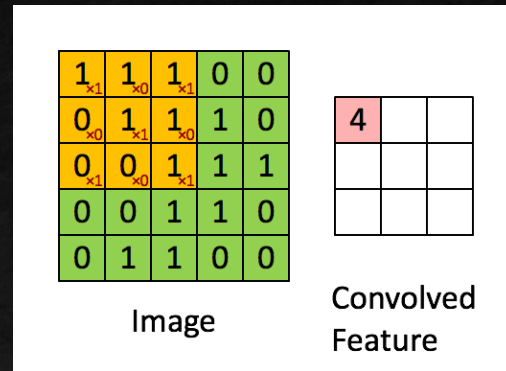
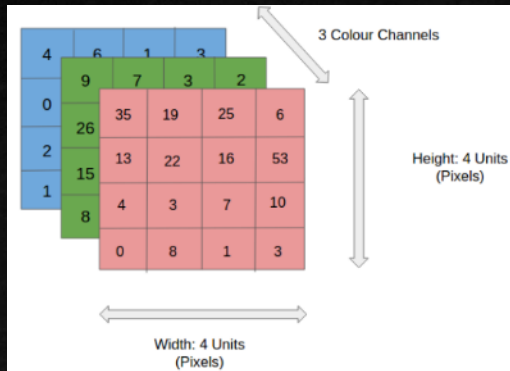


# Why salt body identification?

- Several areas of Earth with large accumulations of oil and gas *also* have huge deposits of salt below the surface.
- But unfortunately, knowing where large salt deposits are precisely is very difficult. Professional seismic imaging still requires expert human interpretation of salt bodies. This leads to very subjective, highly variable renderings. More alarmingly, it leads to potentially dangerous situations for oil and gas company drillers.
- Salts create challenges for seismic imaging, earth modeling, drilling, production, and field development in general.
- Salt body delineation helps us mitigate risks, improve image processing, and plan out field development.



# Quick CNN Overview



Input Layers: Ingest images as multi-dimensional arrays. A 2d colored image will have 3 dimensions:

- Pixel Width
- Pixel Height
- RGB Color Channels

Convolutional Operation: Extracts high-level feature such as edges and unique spatially dependent features.

- Convolutional kernel/filter
- Performs effective feature engineering

Maxpooling: Reduces spatial size of convolved feature by selecting and equalizing feature map.

- Acts as noise suppressant
- Improves computational efficiency
- Max pooling tends to outperform average pooling

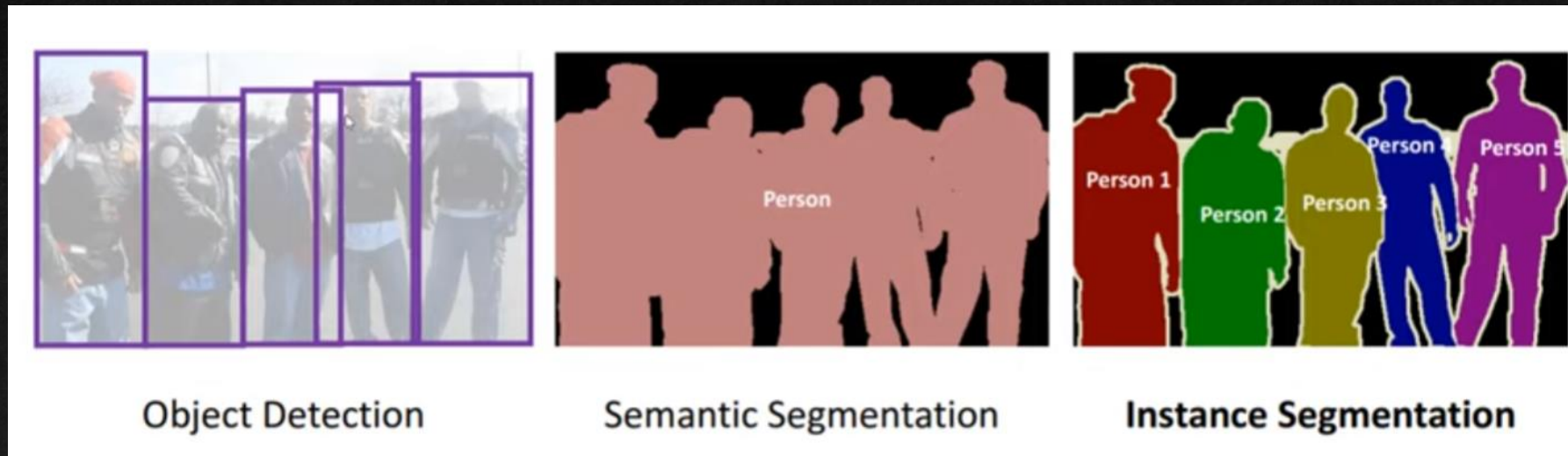
Fully Connected and Output: We pivot back to FFNN layers by flattening CNN output to learn the non-linear relationships between inputs/targets for classification or regression results.

- Model learns how to distinguish between domination and low-level features
- We typically use something like softmax classification but there are others.
- This component can vary based on the target application



# Image Segmentation with U-Net

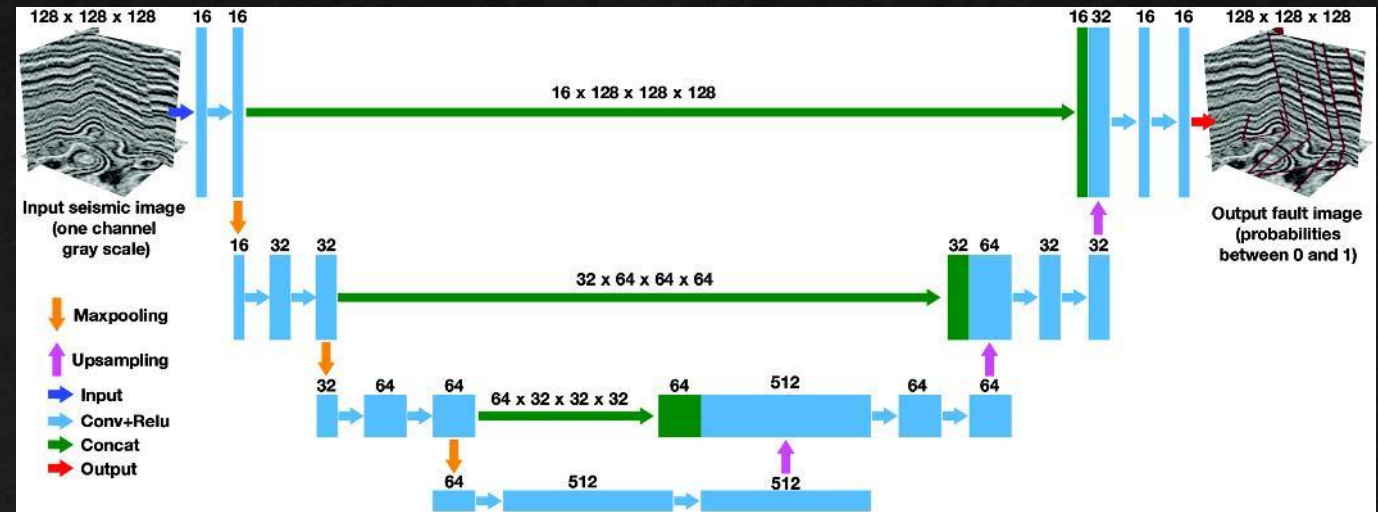
- U-net is a special type of architecture for image segmentation. U-net is a special case of a convolutional neural network where the architecture is configured so that the result is semantically segmented image.
- Semantic Segmentation is basically where you get a binary classification of pixels, yielding a type of image mask. In our case, we are after two classifications salt and no salt.





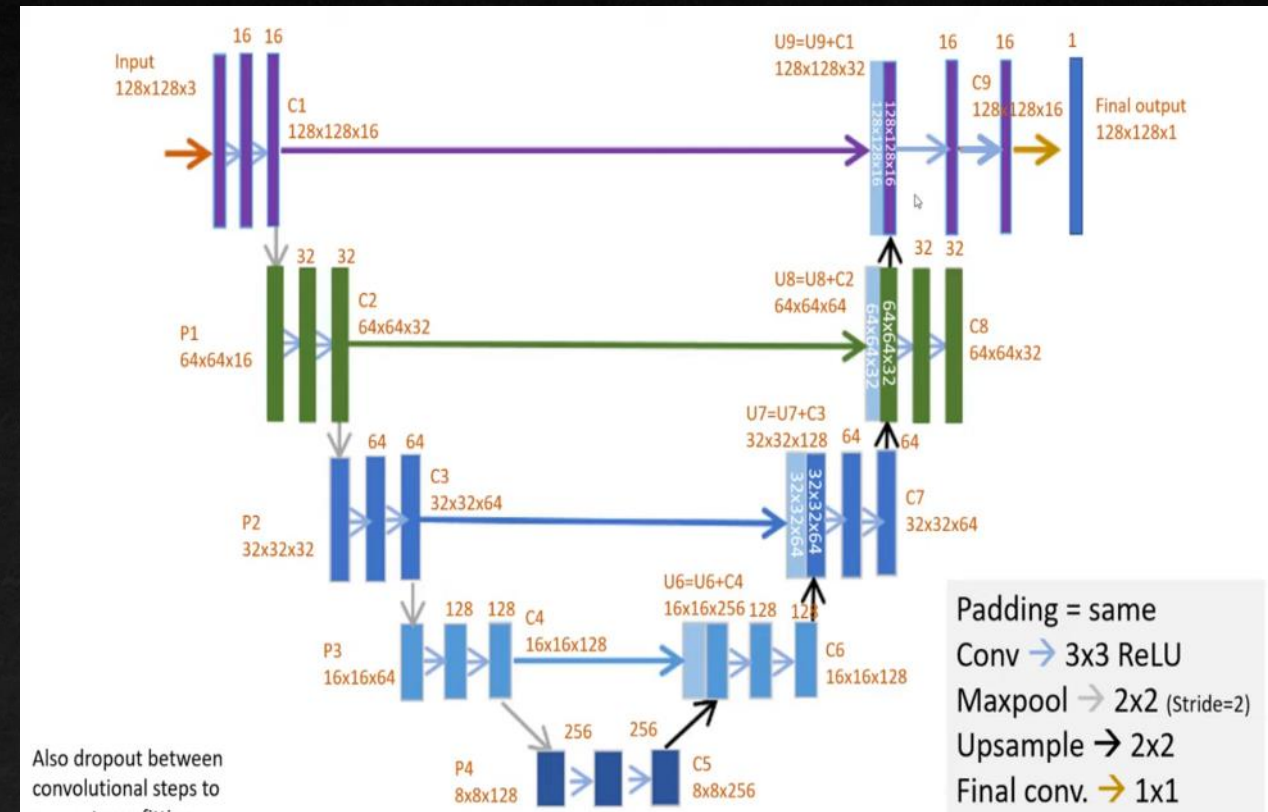
# More on U-Net Differentiators

- U-Net paper: U-Net: Convolutional Networks for Biomedical Image Segmentation by Olaf Ronneberger, Philipp Fischer, Thomas Brox
- Originally designed for medical imaging, now has thousands of applications.
- 2 components:
  - Left: Contraction Path “Encoder”
  - Right: Expander Path “Decoder”
- The concatenation of feature maps from the contraction to the expander path is the genius behind the U-Net architecture and is what enables detailed semantic segmentation.



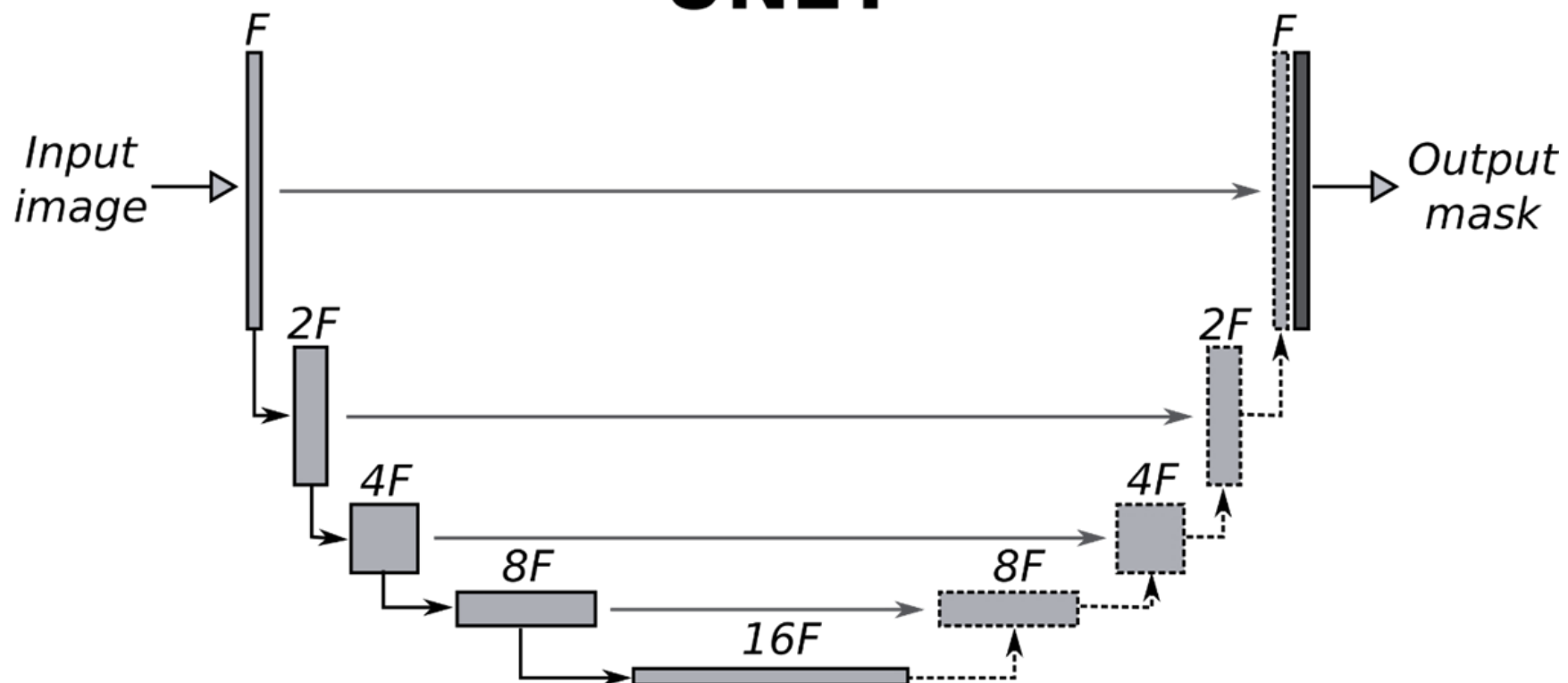
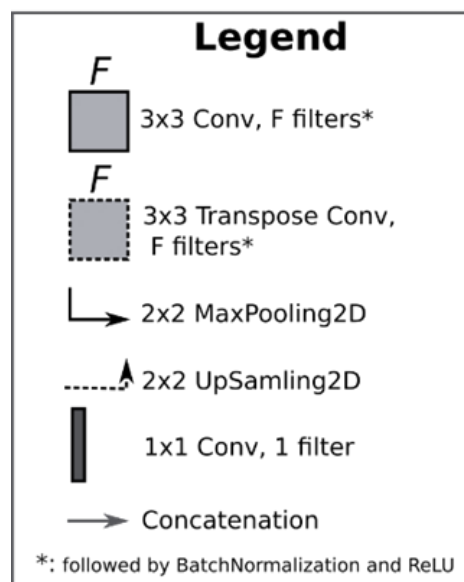
# U-Net Flow Architecture

- The way to interpret a U-Net is to follow the “u” shape and concatenate across.
- Idea here is that you’re adding the corresponding output of your contraction path, to your expansion path.
- Broadly speaking, there are two advantages to using skip connections in this case:
  - They allow gradients to more freely flow through the model, mitigating the issue of vanishing gradients
  - They allow features from the encoder side of the network to the decoder side of the network, adding extra information that might be lost because of the down sampling on the encoder side of the network.



# Another Quick Look

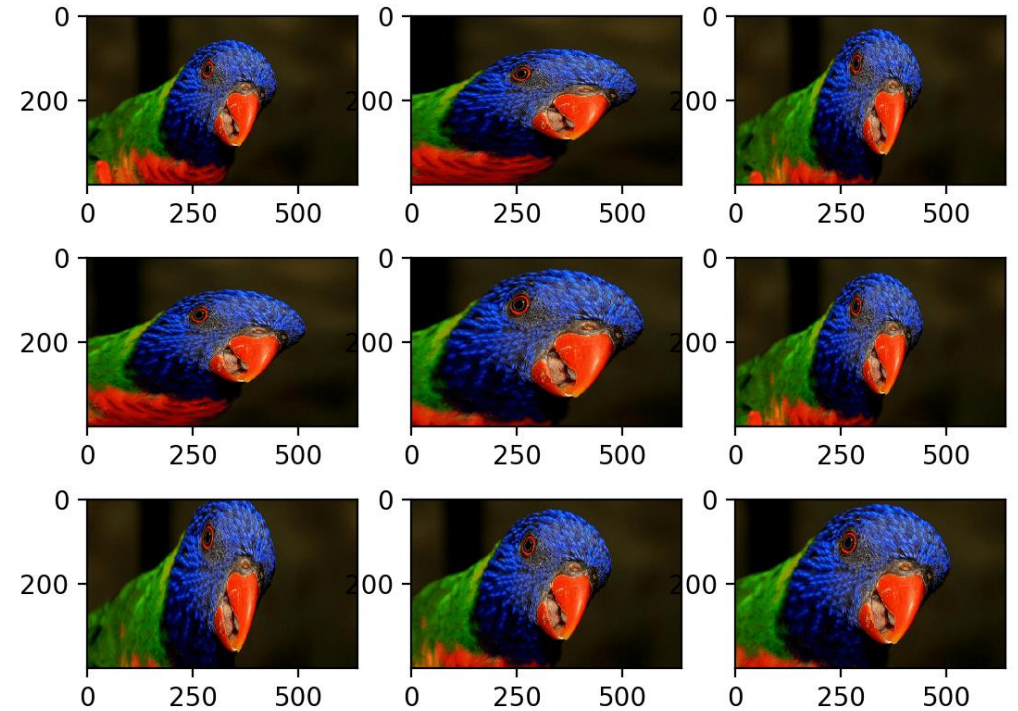
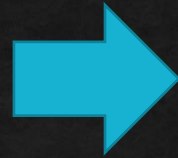
## UNET





# Data Augmentation for CNNs

- Data augmentation helps add to your training dataset by performing various transformations to your existing data inventory.
- In computer vision, this is a cheap and efficient way to boost a model's accuracy on unseen data.





# #SlackHack

## Winning Hack Solution:

- Shout-out as #SlackHack winner on the SIG LinkedIn Page
- Your notebook gets put in the SIG GitHub
- Certificate of Recognition as Winner(s)

*Disclosure: All Kaggle competition data rules still apply to the #slackhack or any use associated with this workshop.*

## #SlackHack

- Immerse yourself in the community of developers
- Grow your portfolio of projects in GitHub or personal website
- Learn new skills with the help of our mentors
- Access to the exclusive Workshop #2 SlackHack Channel
- Challenge: Use the concepts from the course to create your own iteration of salt identification. We will accept other applications of u-net to subsurface data.
- Rules:
  - You must stick to the data provided through the workshop.
  - Work must be your own and must have been developed during the time period of the hack.
  - Hack Deadline: Feb 18th, 2021 (11:59pm CST)
- Judging Criteria:
  - Novelty of Approach
  - Application of Workshop Topics
  - Real World Value of Insights