

Competitive Machine Learning

Kaggle TGS Salt Identification Challenge

Yauhen Babakhin

14 October, 2018

Data Scientist @ Profitero, Minsk, Belarus



WARGAMING.NET
LET'S BATTLE










WARGAMING.NET
LET'S BATTLE

 **Profitero**



WARGAMING.NET
LET'S BATTLE

 **Profitero**

| Competitions Master  | | |
|--|--|--|
| Current Rank 522 of 91,041 | Highest Rank 245 | |
|  2 |  2 |  5 |
| IEEE's Signal Processing S...  · 8 months ago · Top 2% | 9th of 582 | |
| Two Sigma Connect: Renta...  · a year ago · Top 1% | 12th of 2488 | |
| Homesite Quote Conversion  · 3 years ago · Top 3% | 36th of 1764 | |

Part I

Competitive Machine Learning Overview: Process, Types, Benefits/Drawbacks

Part II

Kaggle TGS Salt Identification Challenge

Competitions Platforms

kaggle™



Analytics Vidhya
Learn Everything About Analytics



topcoder™



DRIVEN DATA



Boosters.pro



Competitions Problems

- Classic ML (table data, time series)
- Deep Learning (text, images)



dmlc
XGBoost



- Submitting Predictions File



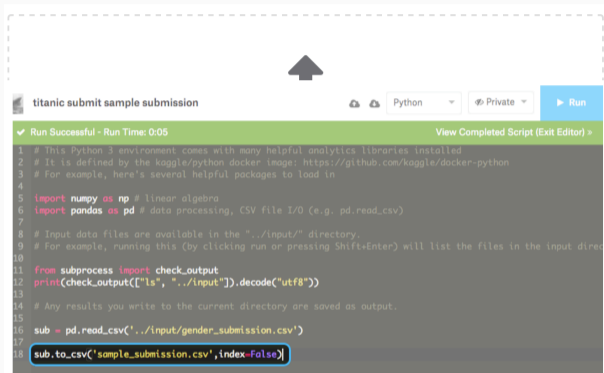
Upload Submission File

File Format

Your submission should be in CSV format. You can upload this in a zip/gz/rar/7z archive, if you prefer.

Competitions Types

- Submitting Predictions File
- Kernels Competitions

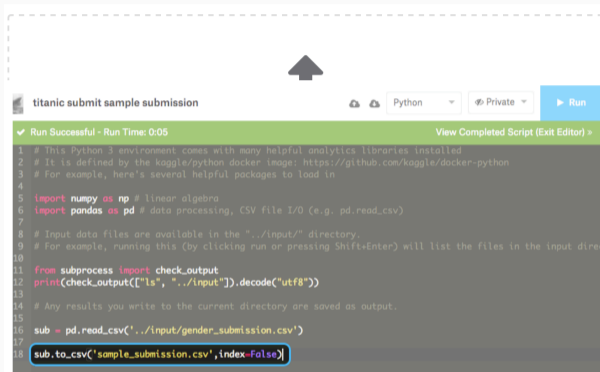


The screenshot shows a Kaggle kernel interface. At the top, there is a title "titanic submit sample submission" and a "Run" button. Below the title, there is a status bar indicating "Run Successful - Run Time: 0:05" and a link to "View Completed Script (Exit Editor)". The main area contains a code editor with the following Python code:

```
1 # This Python 3 environment comes with many helpful analytics libraries installed
2 # It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
3 # For example, here's several helpful packages to load in
4
5 import numpy as np # linear algebra
6 import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
7
8 # Input data files are available in the "../input/" directory.
9 # For example, running this (by clicking run or pressing Shift+Enter) will list the files in the input directory
10
11 from subprocess import check_output
12 print(check_output(["ls", "../input"]).decode("utf8"))
13
14 # Any results you write to the current directory are saved as output.
15
16 sub = pd.read_csv('../input/gender_submission.csv')
17
18 sub.to_csv('sample_submission.csv', index=False)
```

Competitions Types

- Submitting Predictions File
- Kernels Competitions
- Docker Competitions



The screenshot shows a Kaggle kernel editor interface. At the top, there's a title "titanic submit sample submission" and a "Run" button. Below the title, a green bar indicates "Run Successful - Run Time: 0:05". The main area contains a Python script with the following code:

```
1 # This Python 3 environment comes with many helpful analytics libraries installed
2 # It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
3 # For example, here's several helpful packages to load in
4
5 import numpy as np # linear algebra
6 import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
7
8 # Input data files are available in the "../input/" directory.
9 # For example, running this (by clicking run or pressing Shift+Enter) will list the files in the input directory
10
11 from subprocess import check_output
12 print(check_output(["ls", "../input"]).decode("utf8"))
13
14 # Any results you write to the current directory are saved as output.
15
16 sub = pd.read_csv('../input/gender_submission.csv')
17
18 sub.to_csv('sample_submission.csv', index=False)
```



Competitions Benefits



Competitions Benefits for Beginners

1. Playground competitions
2. Work with real data
3. Build initial EDA
4. Build simple models



Competitions Benefits for Juniors

1. Community
2. Get some practical tips&tricks
3. Portfolio projects
4. Interview preparation



Competitions Benefits for Middles

1. New domain/problem type
2. Ideas to apply at the current job
3. Portfolio projects
4. Team work



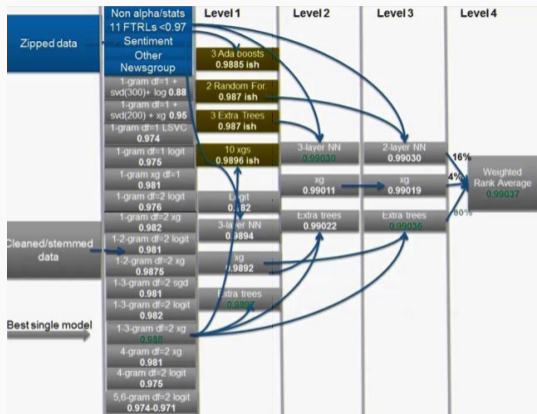
Competitions Benefits for Seniors

1. New domain/problem type
2. Keep up-to-date with the best performing methods
3. Winning prizes
4. Posts and papers



Competitions Drawbacks

1. Not production solutions*
2. Low-quality code
3. Time consuming



http://blog.kaggle.com/2015/12/03/dato-winners-interview-1st-place-mad-professors/

Kaggle TGS Salt Identification Challenge

My Background

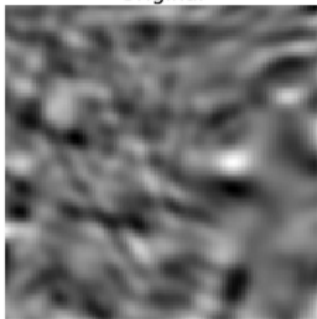
| # | △1w | Team Name | Kernel | Team Members | Score | Entries | Last |
|---|-----|--------------------------------|--------|--------------|--------------------|---------|------|
| 1 | ▲2 | Ding Han Renan Kent AI Lab | | +4 | 0.879 | 384 | 28m |
| 2 | ▲5 | Giba&Heng | | | 0.878 | 185 | 6h |
| 3 | ▼2 | earhian | | | 0.875 | 183 | 9h |
| 4 | ▼2 | SeuTao | | | 0.875 | 254 | 9h |
| 5 | ▲3 | DISK | | | 0.874 | 273 | 6h |
| 6 | ▼2 | Tim & Alex & Anton | | | 0.872 | 202 | 2h |
| 7 | ▼2 | bestfitting | | | 0.870 | 138 | 17m |
| 8 | ▼2 | ZZZ | | | 0.870 | 108 | 2d |
| 9 | ▲5 | b.e.s. | | | 0.870 | 111 | 12h |
| Your Best Entry | | | | | | | |
| Your submission scored 0.870, which is an improvement of your previous score of 0.867. Great job! | | | | | Tweet this! | | |
| 10 | — | [ods.ai] topcoders | | | 0.868 | 204 | 3h |

Typical Competition Pipeline (Exploration)

- Problem Statement
- EDA and Basic Solution Kernels
- Google for Unknown Seed Words
- Look at Top Solutions of Similar Past Competitions

Problem Statement

Original



"GroundTruth" Mask



$$IoU(A, B) = \frac{A \cap B}{A \cup B}$$

EDA and Basic Solution in Kernels

598



Intro to seismic, salt, and how to geophysics

2mo ago 0.62 geophysics, tutorial, beginner, deep learning, object segmentation



Py

104

283



U-net, dropout, augmentation, stratification

2mo ago 0.71 image data, object segmentation



Py

55

163



Explanation of Scoring Metric

2mo ago tutorial, beginner





Py

16



- semantic segmentation
- iou
- unet
- augmentation
- tta
- ...

Look at Top Solutions of Similar Past Competitions

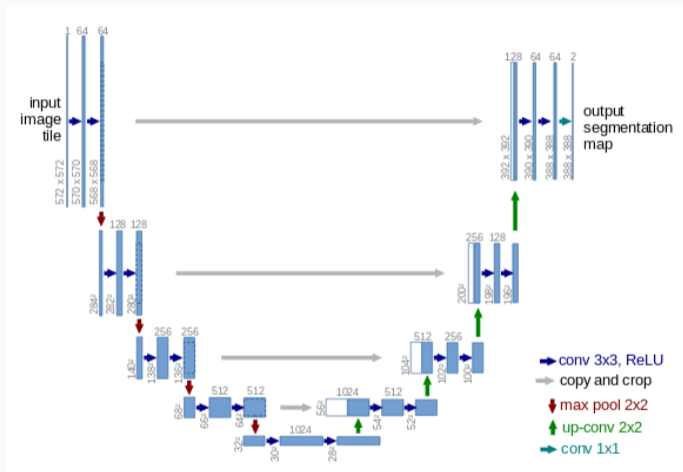
| Segmentation | Competition | Deadline | Type | Evaluation | Reward | Solutions |
|---|--|------------|--|---|----------|---|
|  | 2018 Data Science Bowl Find the nuclei in divergent images to advance medical discovery | 2018-04-17 | Computer vision, Image segmentation | IntersectionOverUnionObjectSegmentation | \$100000 | Forum <ul style="list-style-type: none">• 2nd place solution• 3rd place solution |
|  | Carvana Image Masking Challenge Automatically identify the boundaries of the car in an image | 2017-09-28 | Semantic segmentation, Image processing / Vision | Dice coefficient | \$25000 | Forum <ul style="list-style-type: none">• 1st place solution• 3rd place solution |

<http://ndres.me/kaggle-past-solutions/>

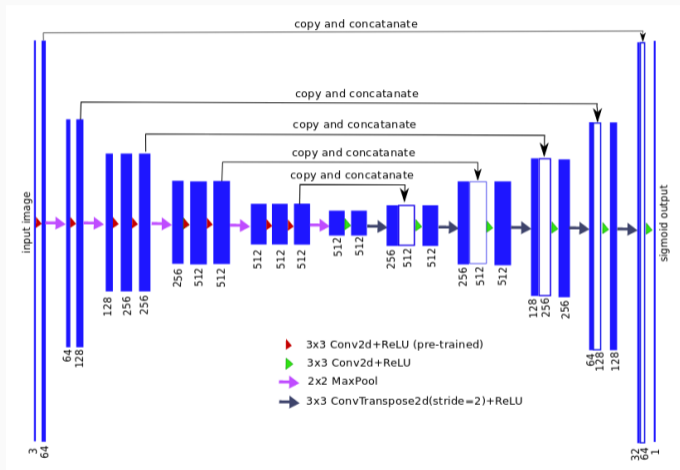
Typical Competition Pipeline (Solution)

- Establish Validation
- Copy/Create Simple Baseline. Fix Bugs
- Read/Invent New Features. Try them
- ...
- Read/Invent New Features. Try them
- Ensemble the Results

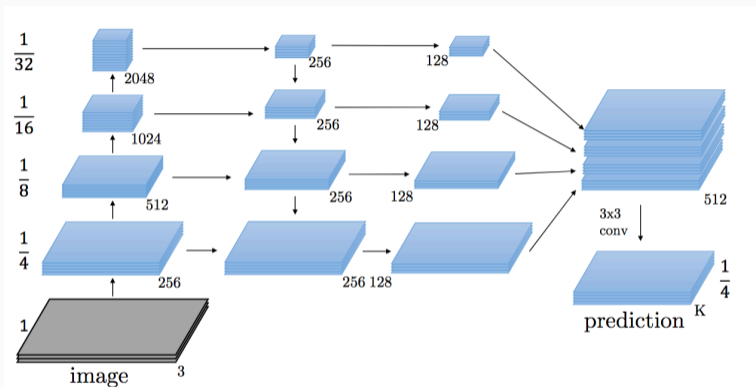
U-Net Architecture



U-Net with Pre-Trained Encoders

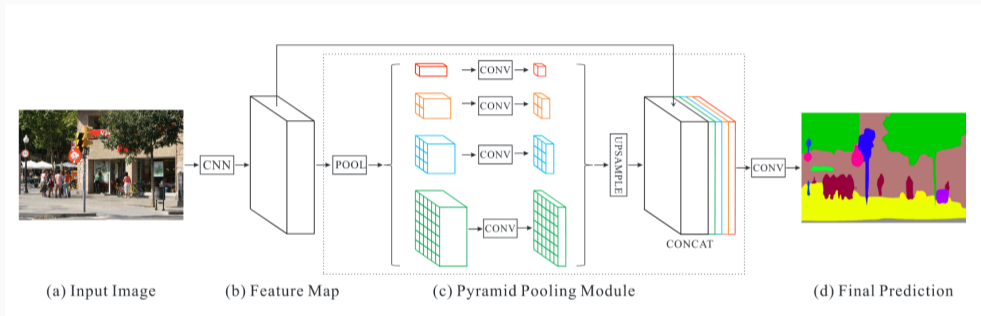


Different Types of Decoders (FPN)



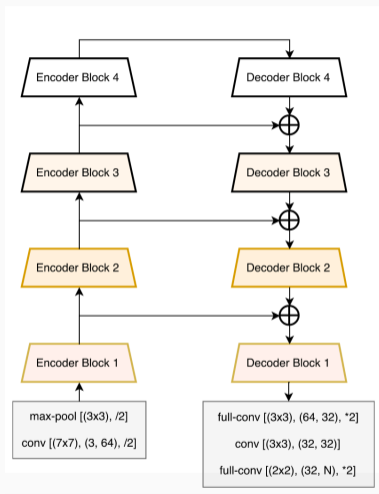
<http://presentations.cocodataset.org/COC017-Stuff-FAIR.pdf>

Different Types of Decoders (PSP)



<https://arxiv.org/pdf/1612.01105.pdf>

Different Types of Decoders (LinkNet)



Keras Segmentation Models Zoo

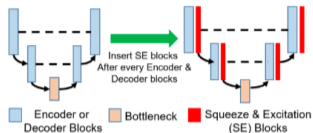
Available models:

- [Unet](#)
- [FPN](#)
- [Linknet](#)
- [PSPNet](#)

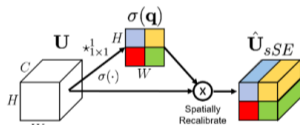
Available backbones:

| Backbone model | Name | Weights |
|----------------|------------------------|---|
| VGG16 | <code>vgg16</code> | <code>imagenet</code> |
| VGG19 | <code>vgg19</code> | <code>imagenet</code> |
| ResNet18 | <code>resnet18</code> | <code>imagenet</code> |
| ResNet34 | <code>resnet34</code> | <code>imagenet</code> |
| ResNet50 | <code>resnet50</code> | <code>imagenet</code> <code>imagenet11k-places365ch</code> |
| ResNet101 | <code>resnet101</code> | <code>imagenet</code> |

Concurrent Spatial and Channel 'Squeeze & Excitation'

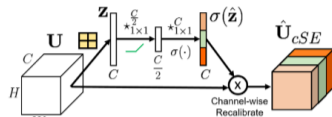


(a) SE block in F-CNN

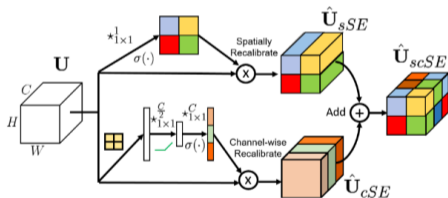


(c) Channel Squeeze and Spatial Excitation (sSE)

$\star_{m \times n}^p$ Convolution with $m \times n$ kernel p channels
 ReLU Global Pooling $\sigma(\cdot)$ Sigmoid

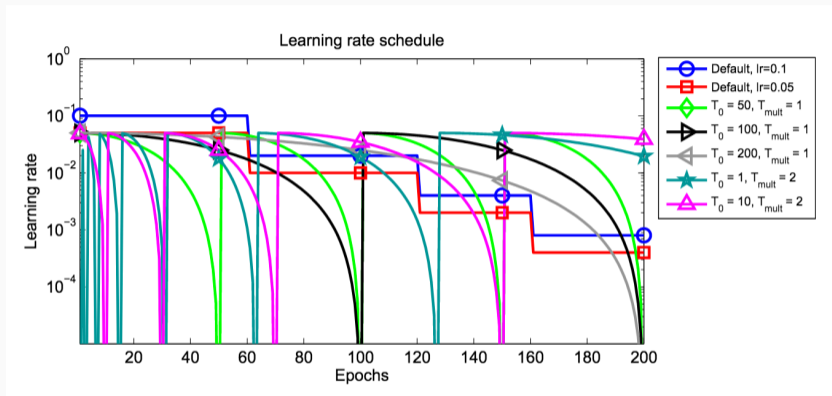


(b) Spatial Squeeze and Channel Excitation (cSE)



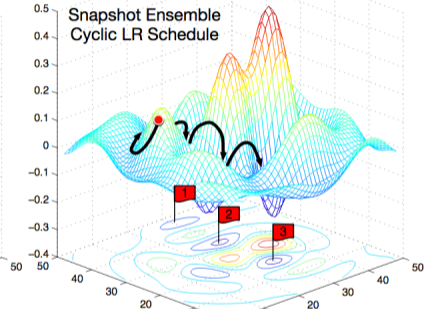
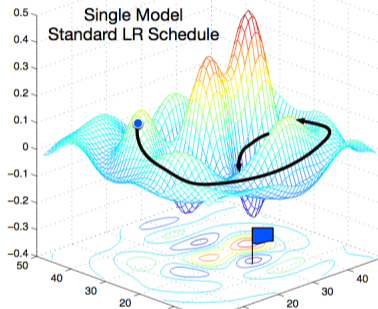
(d) Concurrent Spatial and Channel Squeeze and Channel Excitation (scSE)

Cosine Annealing



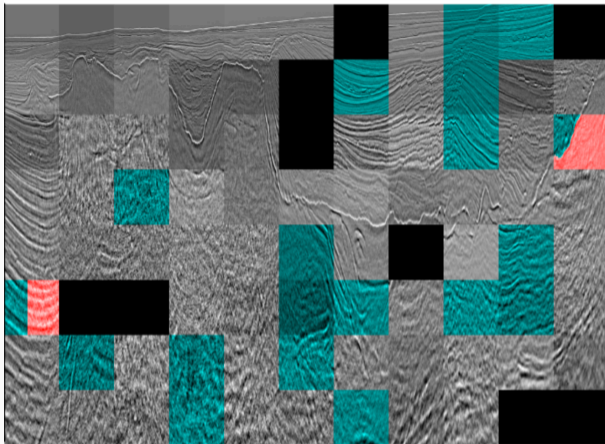
<https://arxiv.org/pdf/1608.03983.pdf>

Snapshot Ensembles



<https://arxiv.org/pdf/1704.00109.pdf>

Bonus: Mosaic



<https://www.kaggle.com/c/tgs-salt-identification-challenge/discussion/66940>

Thank you!

Questions?

 [linkedin.com/in/yauhenbabakhin](https://www.linkedin.com/in/yauhenbabakhin)

 y.babakhin@gmail.com

 [ods.ai: @b.e.s.](#)