

# Special Topics in Applications (AIL861)

## Artificial Intelligence for Earth Observation

### Lecture 24

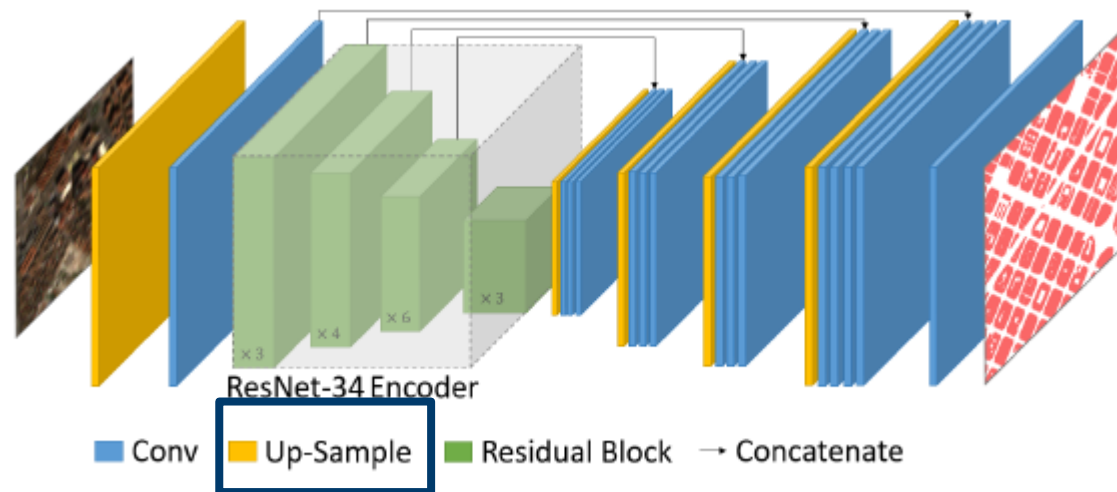
Instructor: Sudipan Saha

# Building Detection from VHR Images

- ✓ What type of problem?
- ✓ Which architectures?

# Building Detection from Sentinel-2 Images

- ✓ Issues with coarser resolution
- ✓ Super-resolution helps



Pushing the limits of Sentinel-2 for building footprint extraction, 2022

# Co-registration Issues

- ✓ Remember our discussion about OSM
- ✓ What if the co-registration/alignment was near-perfect?

# Urban Green Space

- ✓ Very important environmental parameter
- ✓ How to model it?

# Urban Green Space

**Table 1: Performance of U-Net model from base level**

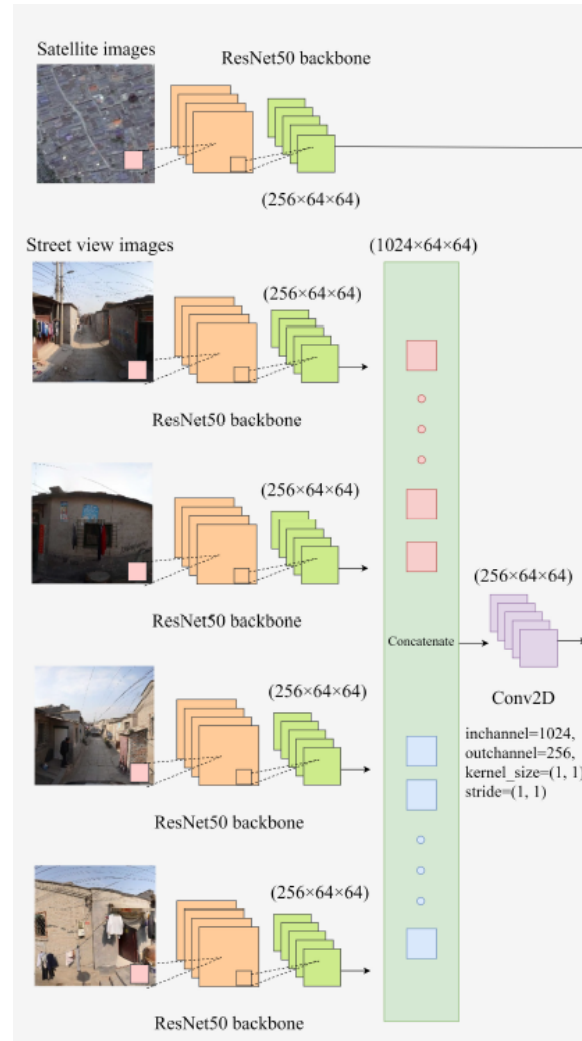
Combination of bands	OA	IoU	F-score	AUC
Red-Green-Blue	0.8329	0.5394	0.5438	0.6894
Red-Green-NIR	0.8531	0.6425	0.6227	0.7373
NDVI-Red-NIR	0.8724	0.6429	0.6921	0.7745
→ NDWI-Red-NIR	0.8977	0.7185	0.7819	0.8457
NDBI-Red-NIR	0.8498	0.6072	0.6583	0.7599
NDVI-NDWI-NDBI	0.8515	0.5846	0.6126	0.7252

Automatically Mapping Urban Green Space Using Sentinel-2 Imagery and Deep Learning Methods in Multiple Cities Worldwide: A Convolutional Neural Network Approach

# Urban Village Classification

- ✓ Urban villages refer to high-density population urban informal settlements characterized by low standards and poverty.
- ✓ Can be treated as semantic segmentation problem.
- ✓ Semantic segmentation using only satellite images or can we use some additional data source?

# Combining Satellite and Street-view Images





# Street View Images

	Advantages	Disadvantages
SVI	<ul style="list-style-type: none"> <li>- High resolution and detail</li> <li>- Represent more closely how environments are perceived and experienced by people</li> <li>- Street view images can provide ground-level details that aerial images lack</li> <li>- Relatively low cost (i.e. freely available data thanks to both commercial and crowdsourced services)</li> <li>- Some objects in area with a dense SVI coverage may be observed more than once</li> </ul>	<ul style="list-style-type: none"> <li>- Coverage is limited, only area near streets are imaged</li> <li>- Imbalanced data coverage (both spatial and temporal)</li> <li>- Limited temporal coverage and revisit periods (i.e. sometimes once in a decade)</li> <li>- Difficulty in expanding research scale due to extensive processing time</li> <li>- Complicated processing workflow, advanced computer vision techniques are required to process the data</li> <li>- Tall objects (buildings, trees, etc.) may be only partially observed from SVI</li> <li>- The position and light condition of taking SVI may vary in the same point</li> </ul>
Aerial	<ul style="list-style-type: none"> <li>- Wider spatial coverage</li> <li>- Fine time granularity</li> <li>- Can be used to observe broad trends</li> <li>- Capture the overall information in a large scale</li> <li>- Data volume is relatively small</li> </ul>	<ul style="list-style-type: none"> <li>- Certain mismatch between satellite's top-down viewpoint and human-scale viewpoint</li> <li>- Cannot capture some details in building facade</li> <li>- Difficult for human's with limited experience to interpret</li> <li>- Openly available satellite imagery generally has too coarse spatial resolution</li> </ul>

# Combining Satellite Images and Social Media Texts

- ✓ Sources for social media text
- ✓ Advantages, disadvantages



## TRANSFORMATIVE TECHNOLOGY

# Social Media + Satellite Imagery = Predictive Power

When disaster strikes, rescuers gather massive amounts of information nearly real-time from Twitter. Data scientists are developing a way to fuse the timely social media data with reliable satellite imagery to guide disaster relief, track the spread of disease and more.

August 26, 2019 | 6 min read

# Social Media Text Pre-Processing

- ✓ Filtering data by geotags
- ✓ Text pre-processing, e.g., lowercasing all text, removing blocks of digits.

# GloVe

- ✓ GloVe: Global Vectors for Word Representation.
- ✓ GloVe is an unsupervised learning algorithm for obtaining vector representations for words.  
Training is performed on aggregated global word-word co-occurrence statistics from a corpus.
- ✓ GloVe Twitter: Pre-trained glove vectors based on 2B tweets.

# CrisisMMD Dataset

- ✓ CrisisMMD: Multimodal Crisis Dataset
- ✓ The CrisisMMD multimodal Twitter dataset consists of several thousands of manually annotated tweets and images collected during seven major natural disasters including earthquakes, hurricanes, wildfires, and floods that happened in the year 2017 across different parts of the World.

# CrisisMMD Dataset: Task 1

- Informative
- Not informative

# CrisisMMD Dataset: Task 2

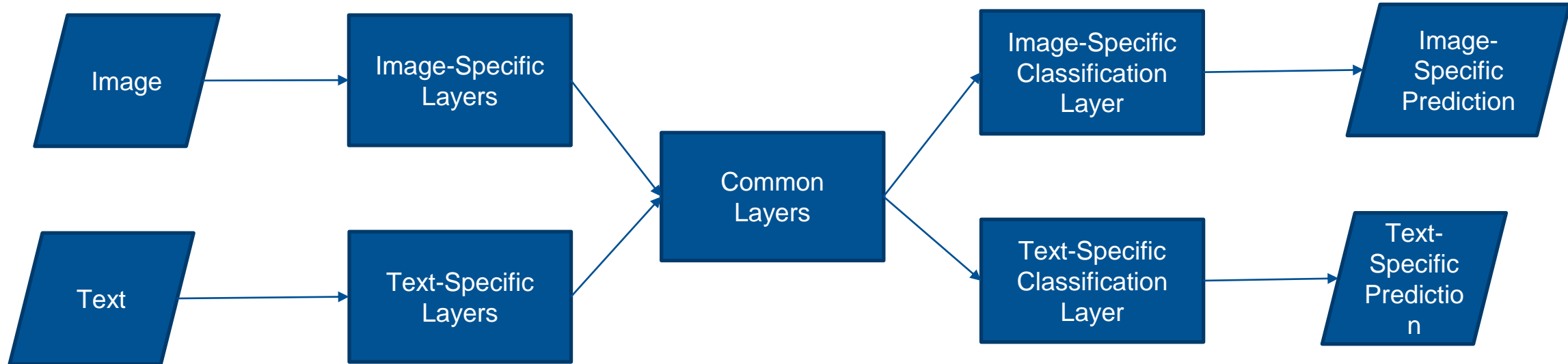
- Affected individuals
- Infrastructure and utility damage
- Injured or dead people
- Missing or found people
- Rescue, volunteering or donation effort
- Vehicle damage
- Other relevant information



# CrisisMMD Dataset: Task 3

- Severe damage
- Mild damage
- Little or no damage
- Don't know or can't judge

# Image + Text Classification



# Image + Text Classification

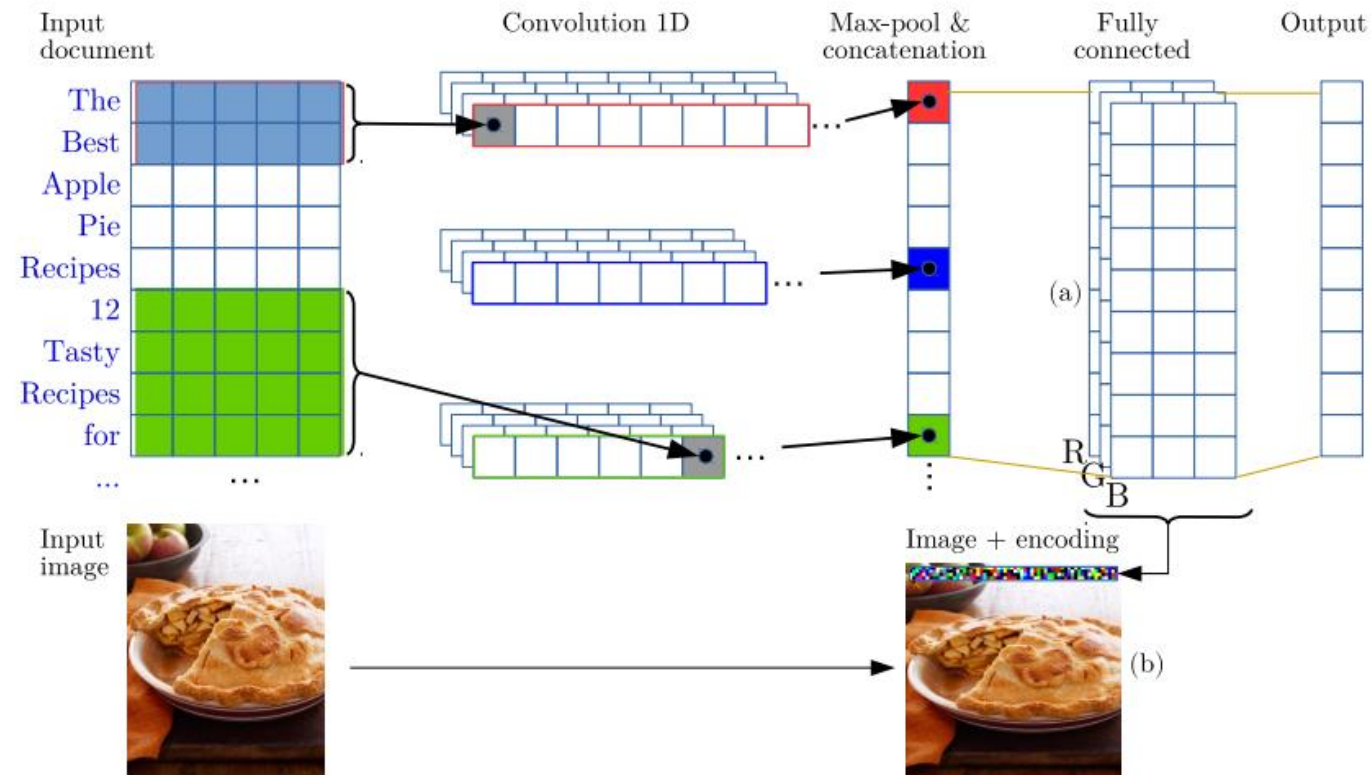


Fig. 2: The proposed text and image fusion model for deep multi-modal classification. The text encoded (a) is passed to the output layer. After the training step only the text features (a) are extracted and then drawn over the original image in order to generate a new multimodal dataset.