Deep Learning Competence Center &
Smart Data and Knowledge Services

Emergency Response with Deep Neural Networks and Satellite Imagery

Benjamin Bischke, Damian Borth

Talk-ID: 23479

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Accessibility of Remote Sensed Imagery

What can be done with it?

Agriculture: Water Utilization

What can be done with it?

Business: Mining Output

What can be done with it?

Business: Mining Output

What can be done with it?

Energy: Pollution Monitoring

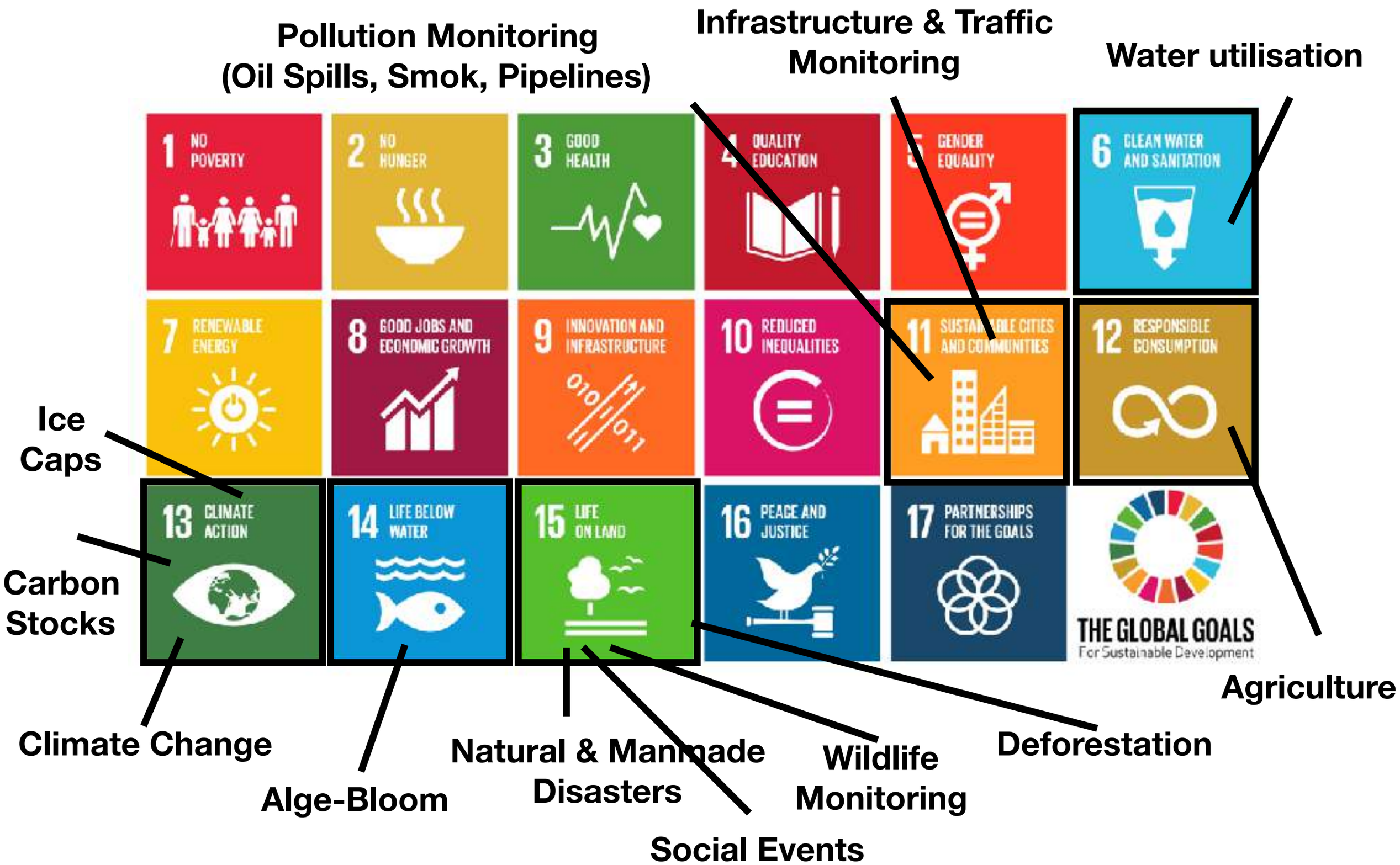
What can be done with it?

What can be done with it?

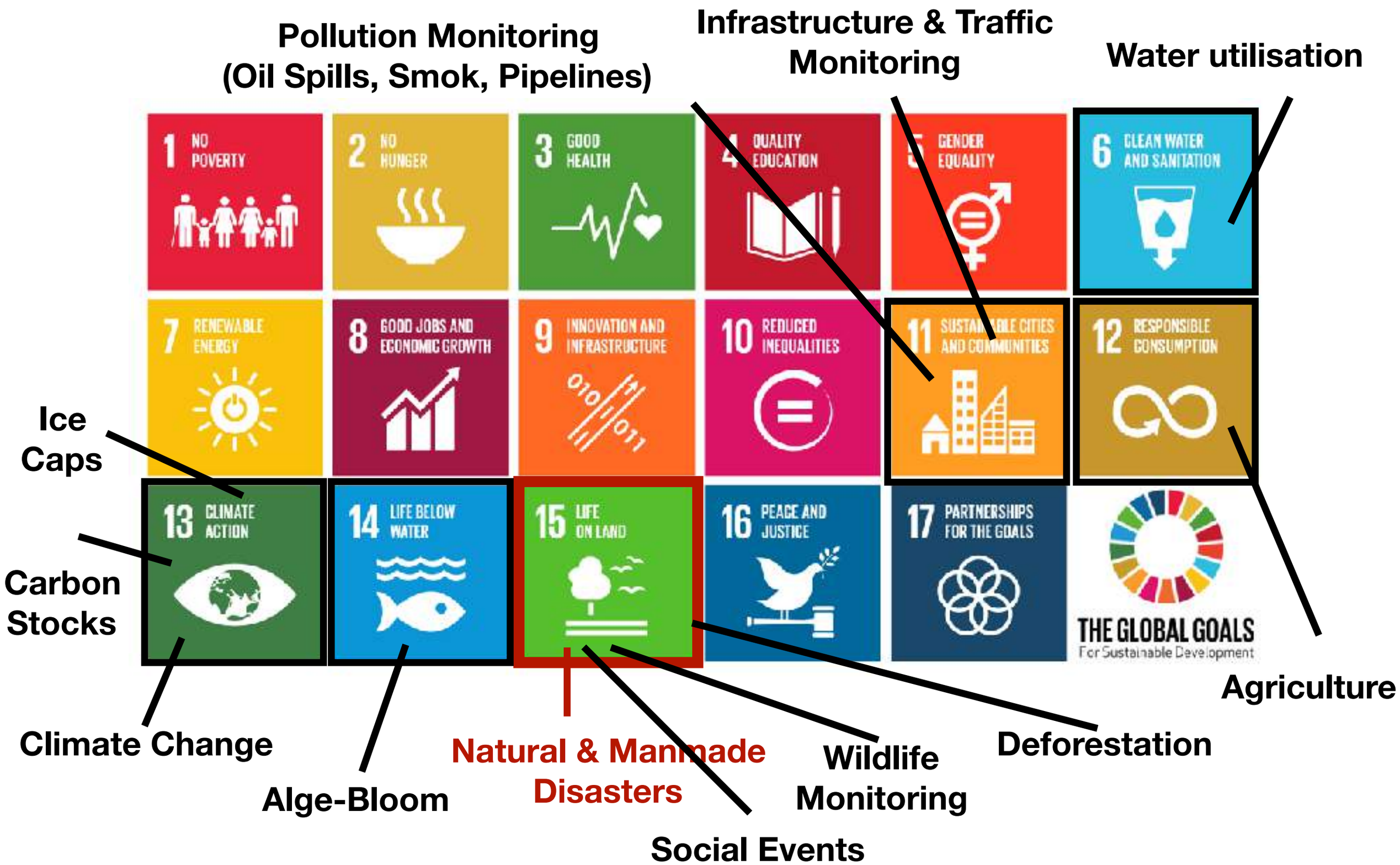
Environment: Deforestation

Satellite Analysis - What can be done with it?

Satellite Analysis - What can be done with it?



Satellite Analysis - What can be done with it?



Situation Awareness with Remote Sensing

Earthquakes

Copernicus Emergency Management Service, Aug. 2016

Wildfires

NASA, Oct. 2007

Flooding

NOAA Remote Sensing Division, Aug. 2016

Dam Break

NASA, Nov. 2015

Situation Awareness with Remote Sensing

Earthquakes

Copernicus Emergency Management Service, Aug. 2016

Wildfires

NASA, Oct. 2007

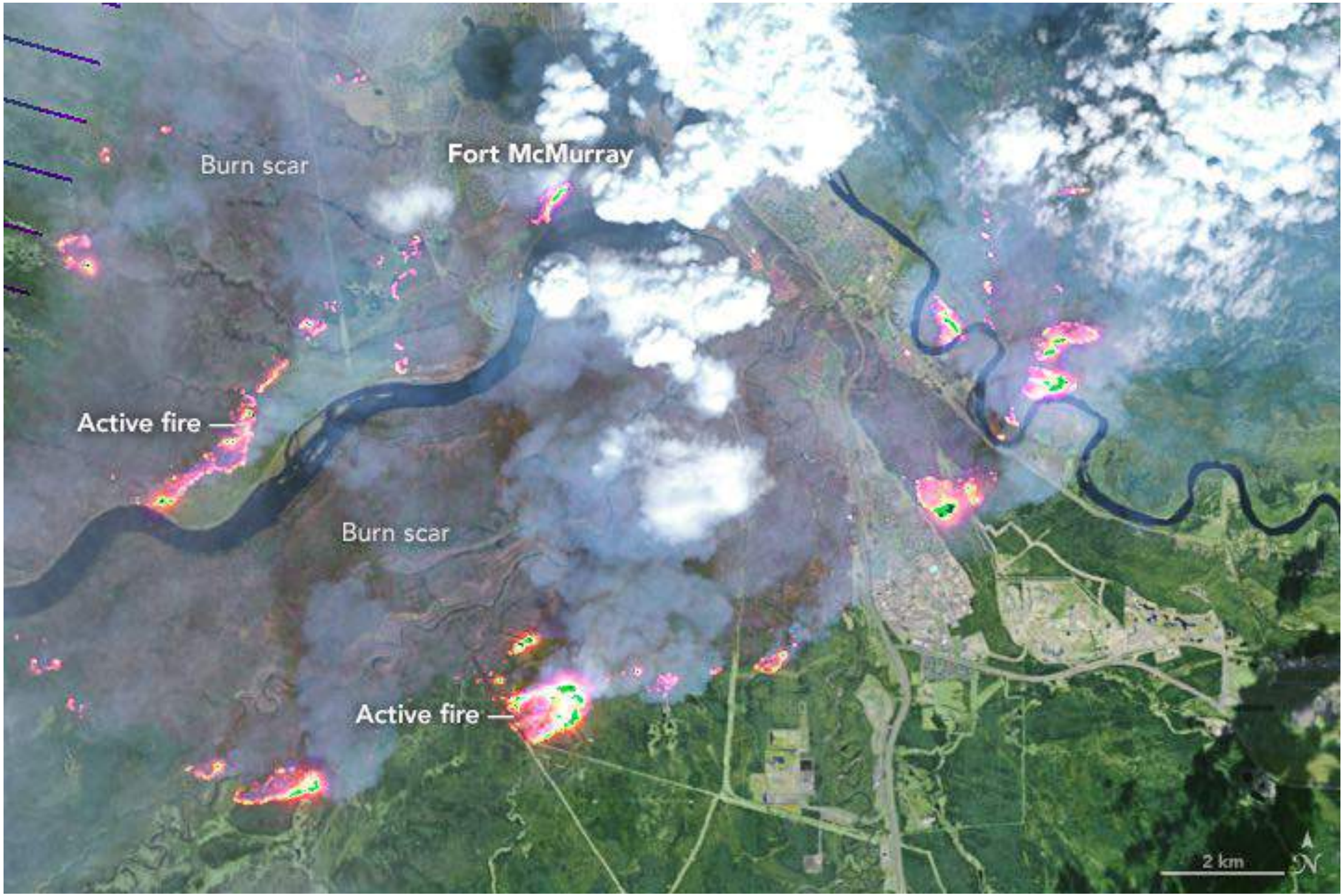
Flooding

NOAA Remote Sensing Division, Aug. 2016

Dam Break

NASA, Nov. 2015

Is Satellite Imagery enough?

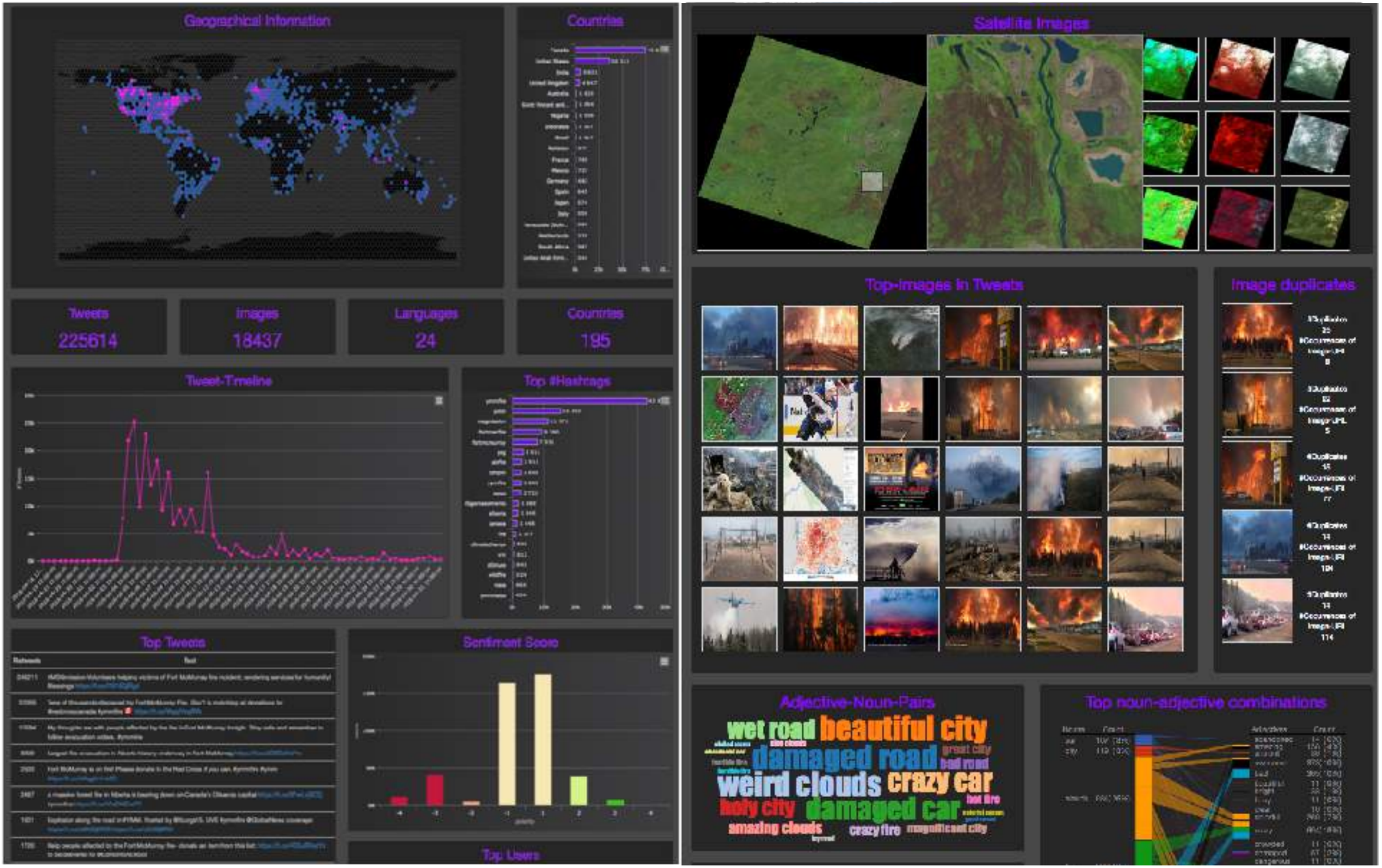


NASA, May 2016

Satellite Image Processing

- Satellite scenes from Landsat8-satellite
- Wildfire only visible in composite of certain bands

“DeepEye” Visualization



Situation Awareness with Remote Sensing

Earthquakes

Copernicus Emergency Management Service, Aug. 2016

Wildfires

NASA, Oct. 2007

Flooding

NOAA Remote Sensing Division, Aug. 2016

Dam Break

NASA, Nov. 2015

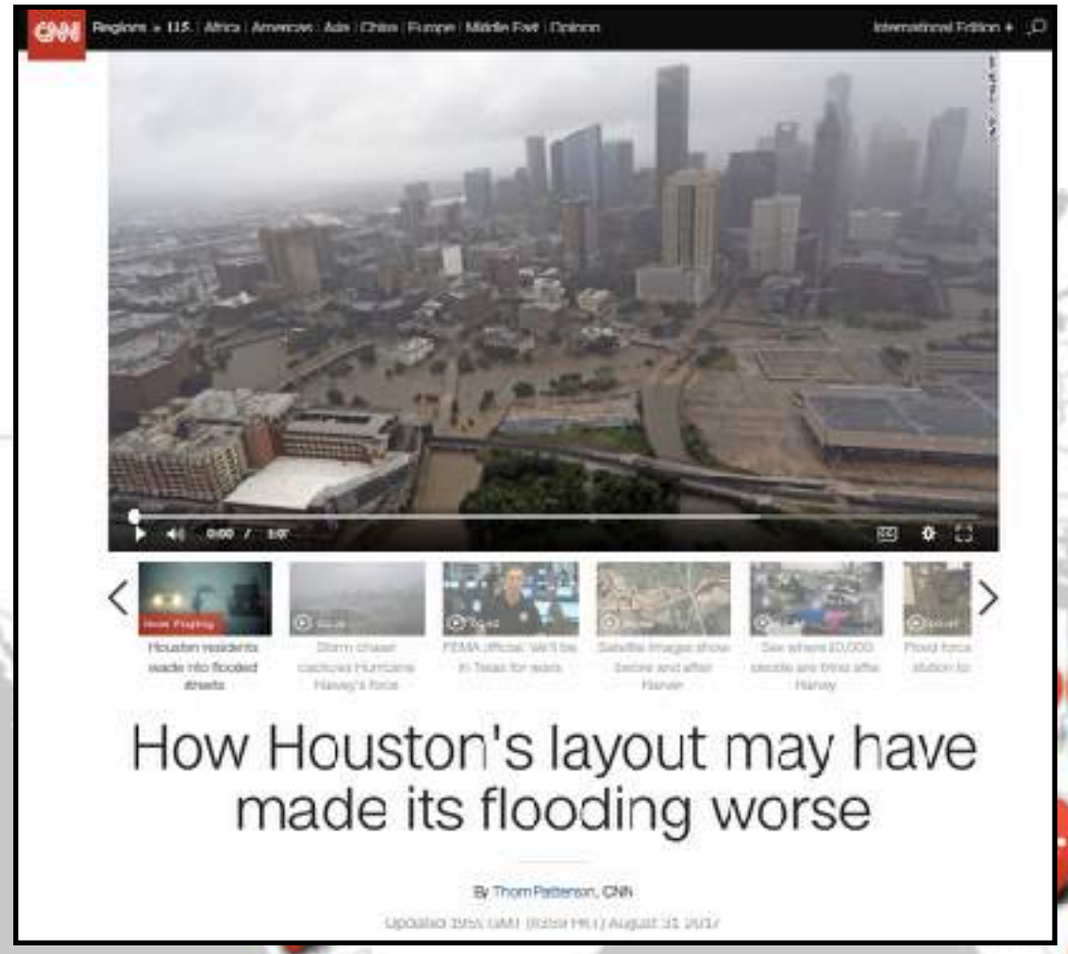
Current Natural Disasters

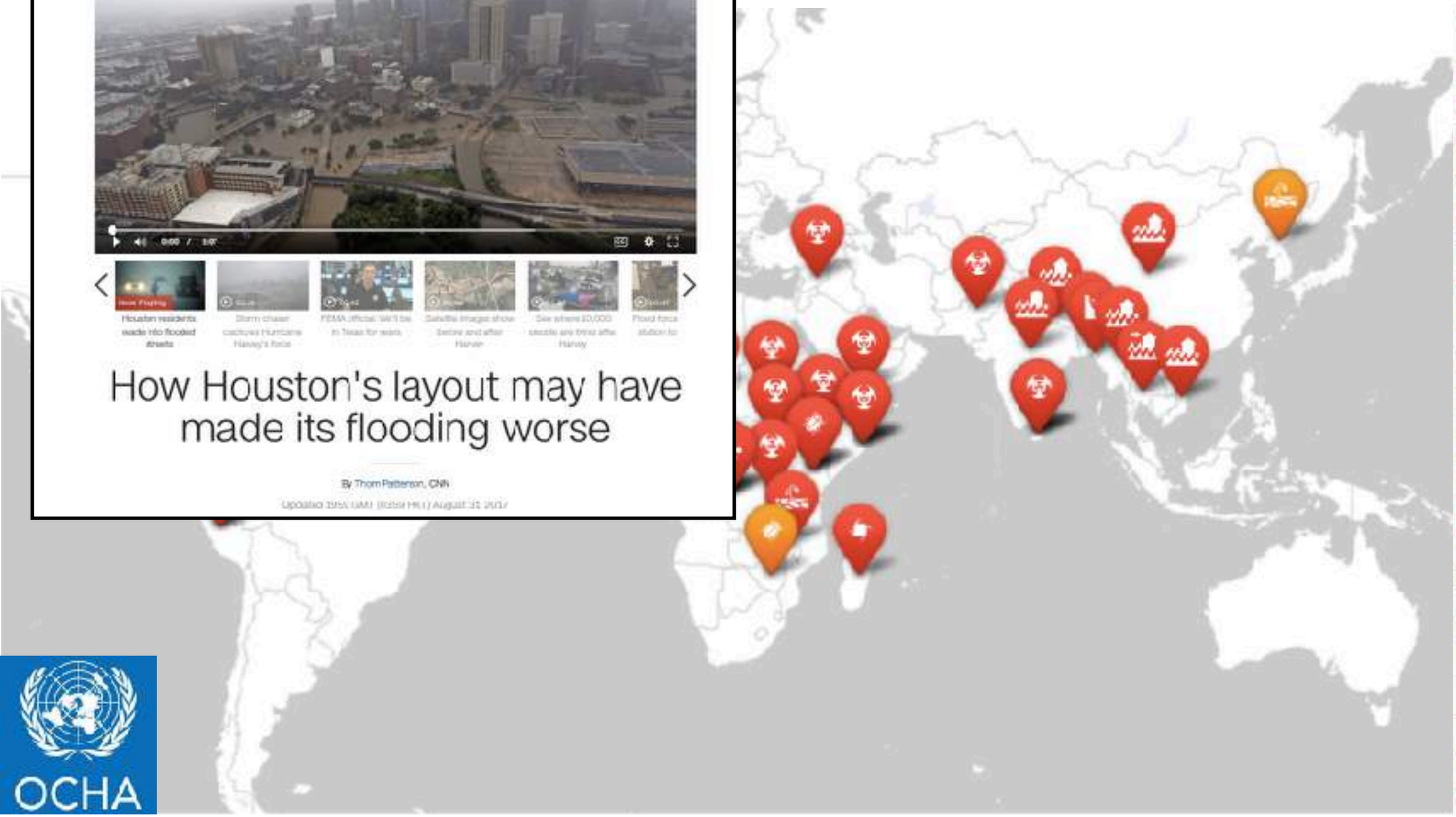
Type

- Cold Wave 73
- Drought 122
- Earthquake 378
- Epidemic 158
- Extratropical Cyclone 19
- Fire 5
- Flash Flood 222
- Flood 1,241
- Heat Wave 6
- Insect Infestation 8
- Land Slide 204
- Mud Slide 39
- Other 11
- Severe Local Storm 141
- Snow Avalanche 19
- Storm Surge 11
- Technological Disaster 42
- Tropical Cyclone 458
- Tsunami 15
- Volcano 102
- Wild Fire 54

Current Natural Disasters

Current Natural Disasters



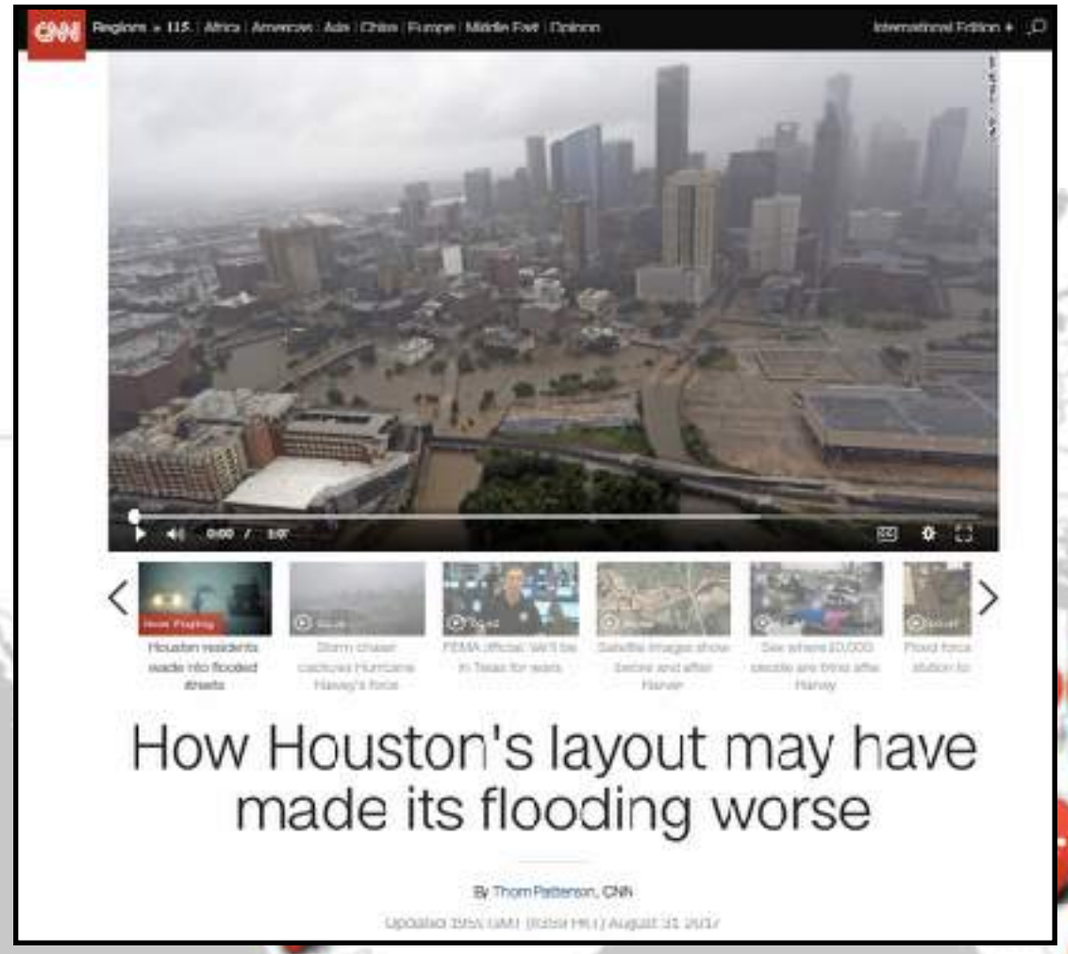


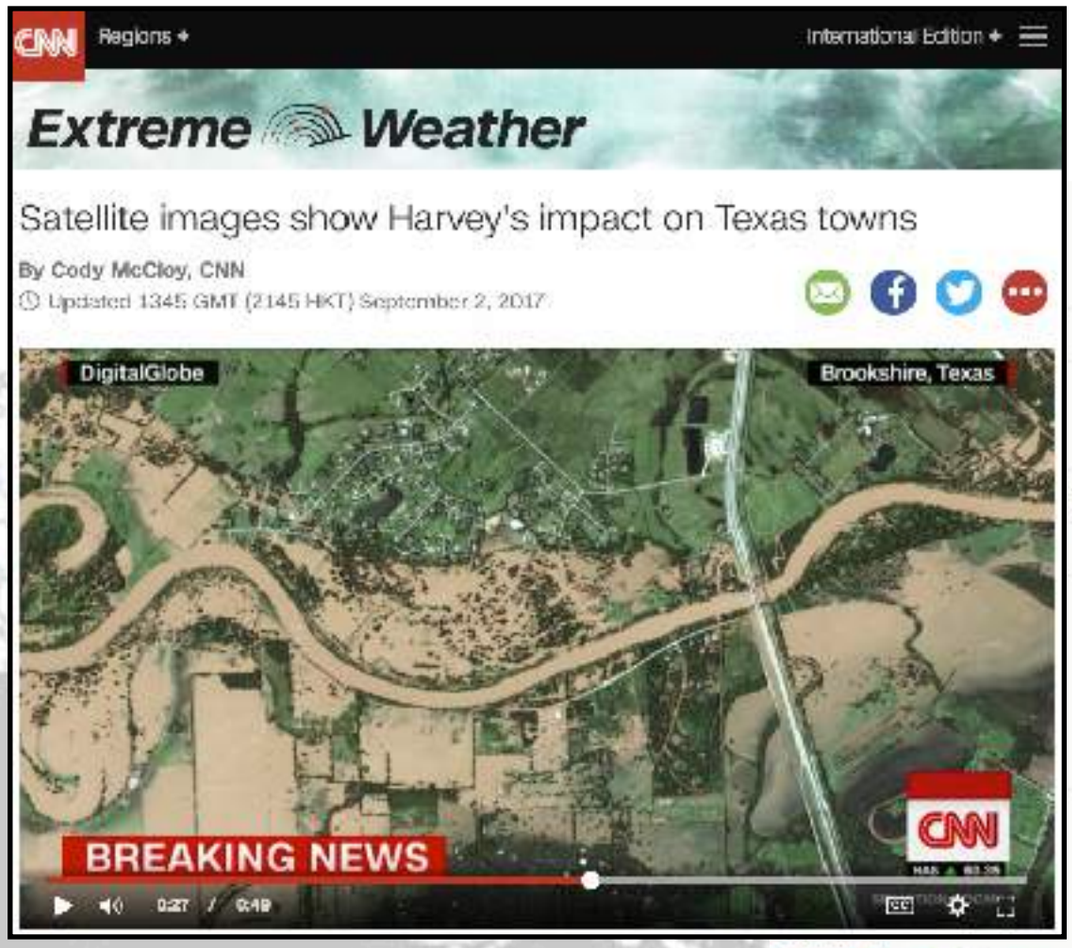
Type

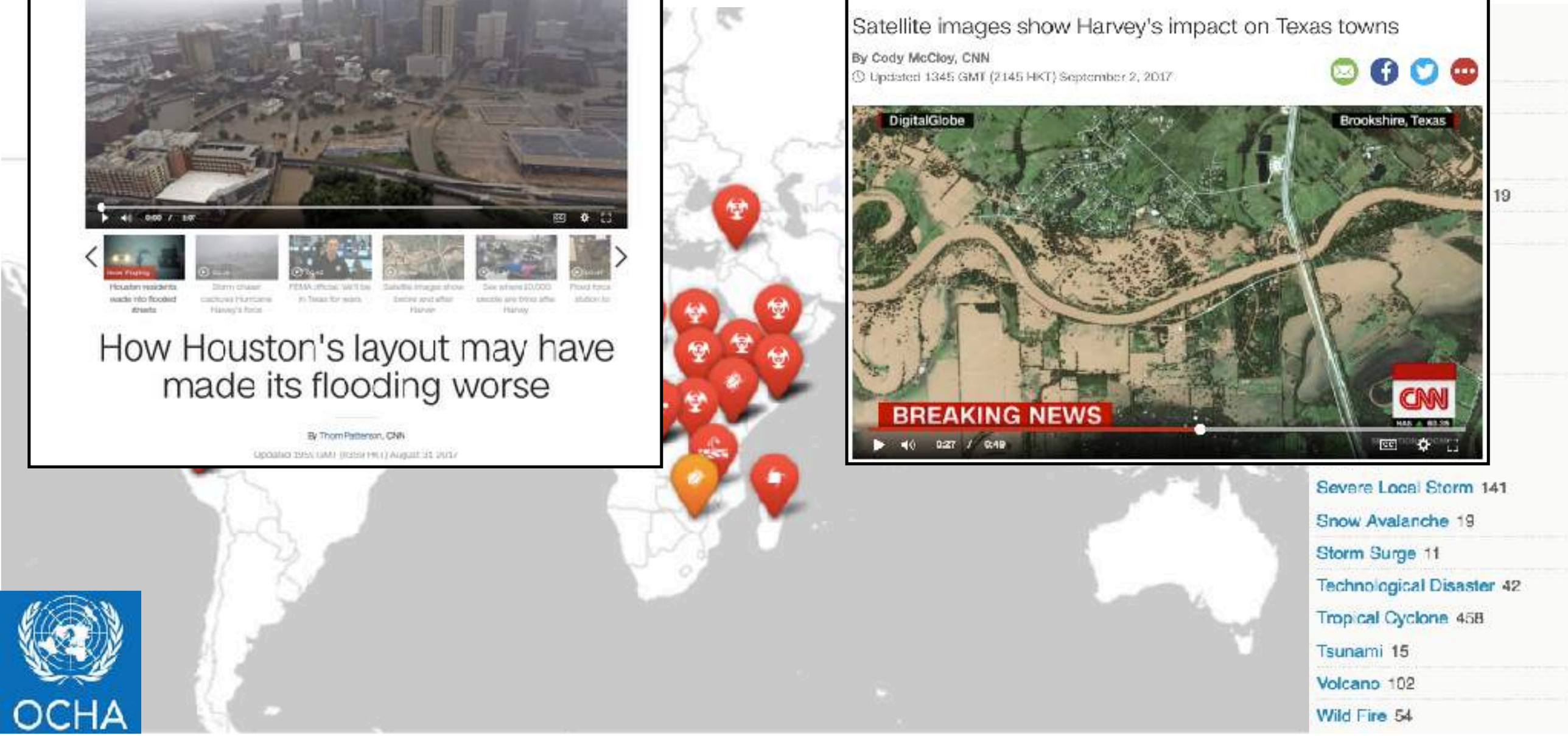
- Cold Wave 73
- Drought 122
- Earthquake 378
- Epidemic 158
- Extratropical Cyclone 19
- Fire 5
- Flash Flood 222
- Flood 1,241
- Heat Wave 6
- Insect Infestation 8
- Land Slide 204
- Mud Slide 39
- Other 11
- Severe Local Storm 141
- Snow Avalanche 19
- Storm Surge 11
- Technological Disaster 42
- Tropical Cyclone 458
- Tsunami 15
- Volcano 102
- Wild Fire 54

OCHA

Current Natural Disasters



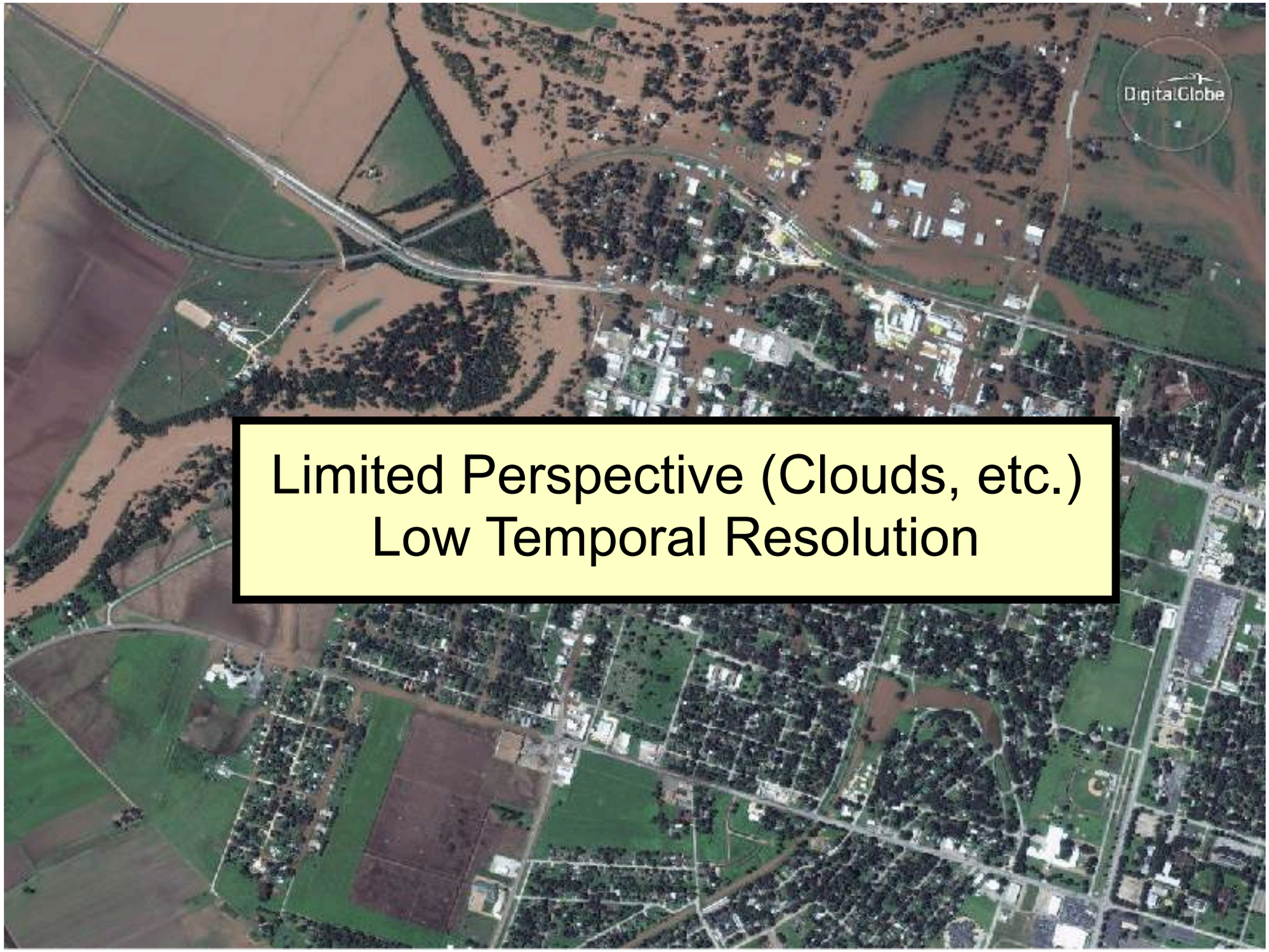




Is Satellite Imagery enough?

DigitalGlobe, October 2017

Is Satellite Imagery enough?



Limited Perspective (Clouds, etc.)
Low Temporal Resolution

DigitalGlobe, October 2017

Contextual Enrichment of Satellite Imagery (Idea)

DigitalGlobe, October 2017

The Multimedia Satellite Task at MediaEval 2017

Multimedia Satellite Task - Overview

- **Goal:** Combine Satellite Imagery with Social Multimedia

- Focus on **Flooding Events**

- Two Subtasks:
  - Disaster Image Retrieval from Social Media (DIRSM)
  - Flood Detection in Satellite Imagery (FDSI)

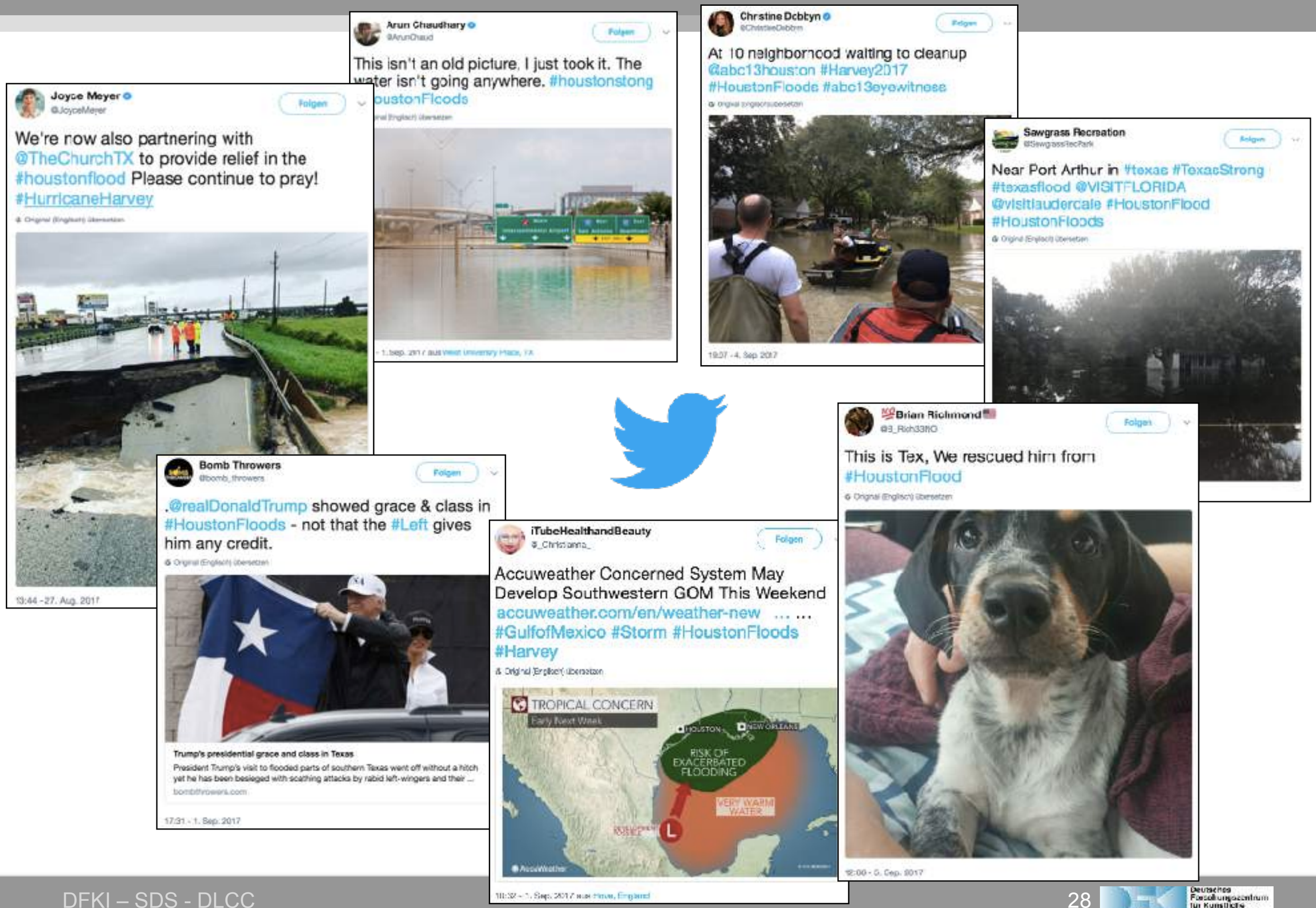
Multimedia Satellite Task - Overview

- **Goal:** Combine Satellite Imagery with Social Multimedia

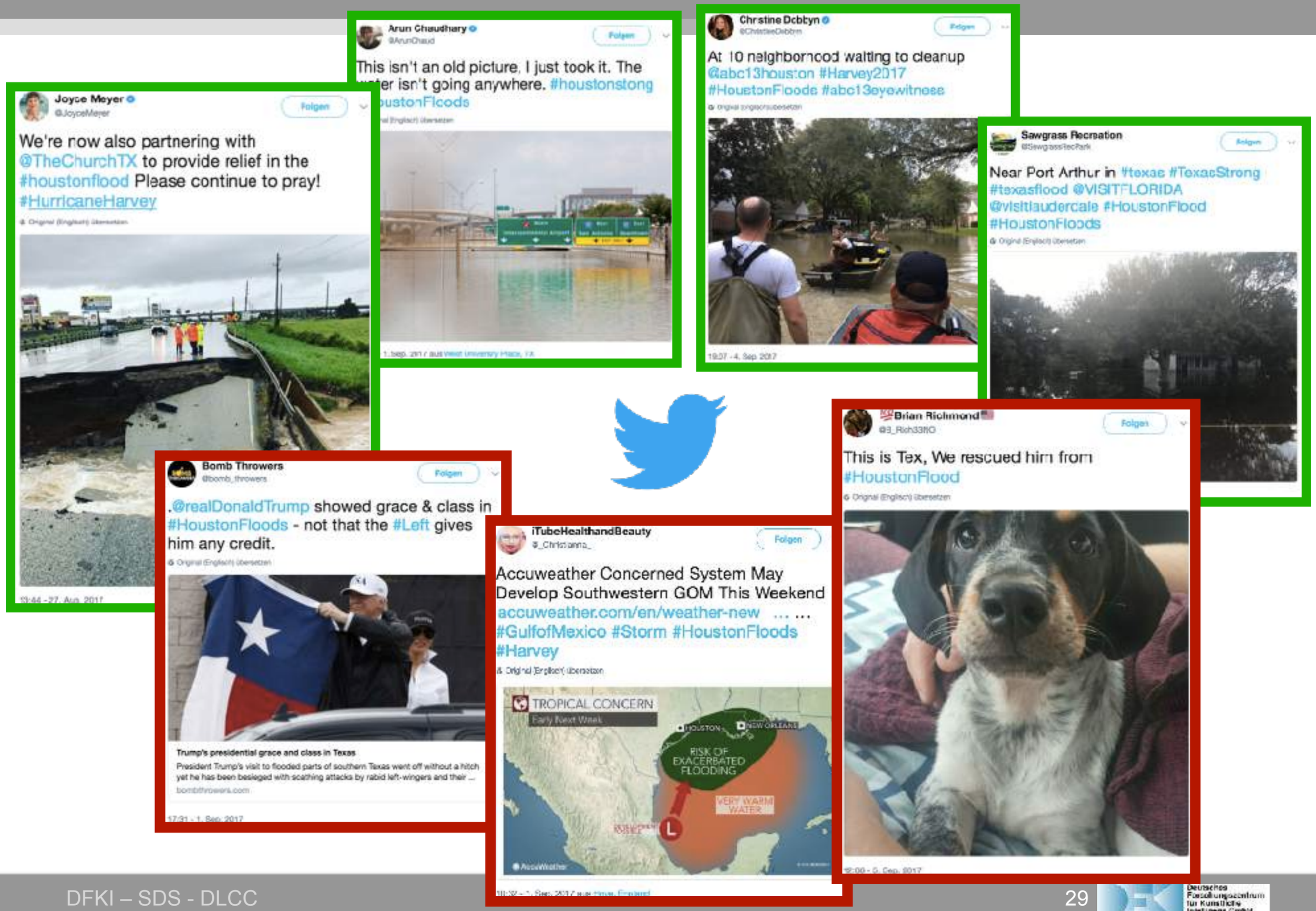
- Focus on **Flooding Events**

- Two Subtasks:
  - **Disaster Image Retrieval from Social Media (DIRSM)**
  - Flood Detection in Satellite Imagery (FDSI)

Disaster Image Retrieval from Social Media



Disaster Image Retrieval from Social Media



Multimedia Satellite Task - Overview

- **Goal:** Combine Satellite Imagery with Social Multimedia

- Focus on **Flooding Events**

- Two Subtasks:
  - Disaster Image Retrieval from Social Media (DIRSM)
  - **Flood Detection in Satellite Imagery (FDSI)**

Flood Detection in Satellite Imagery

Run Submissions and Evaluation

- Up to 5 runs for each subtask:
  - **DIRSM**: three required runs only with the provided dev set (Visual, textual, both)
  - **FDSI**: three required runs only with the provided dev set

- Standard Evaluation Metrics:
  - **DIRSM**: Average Precision at different cutoffs
  - **FDSI**: Intersection over Union

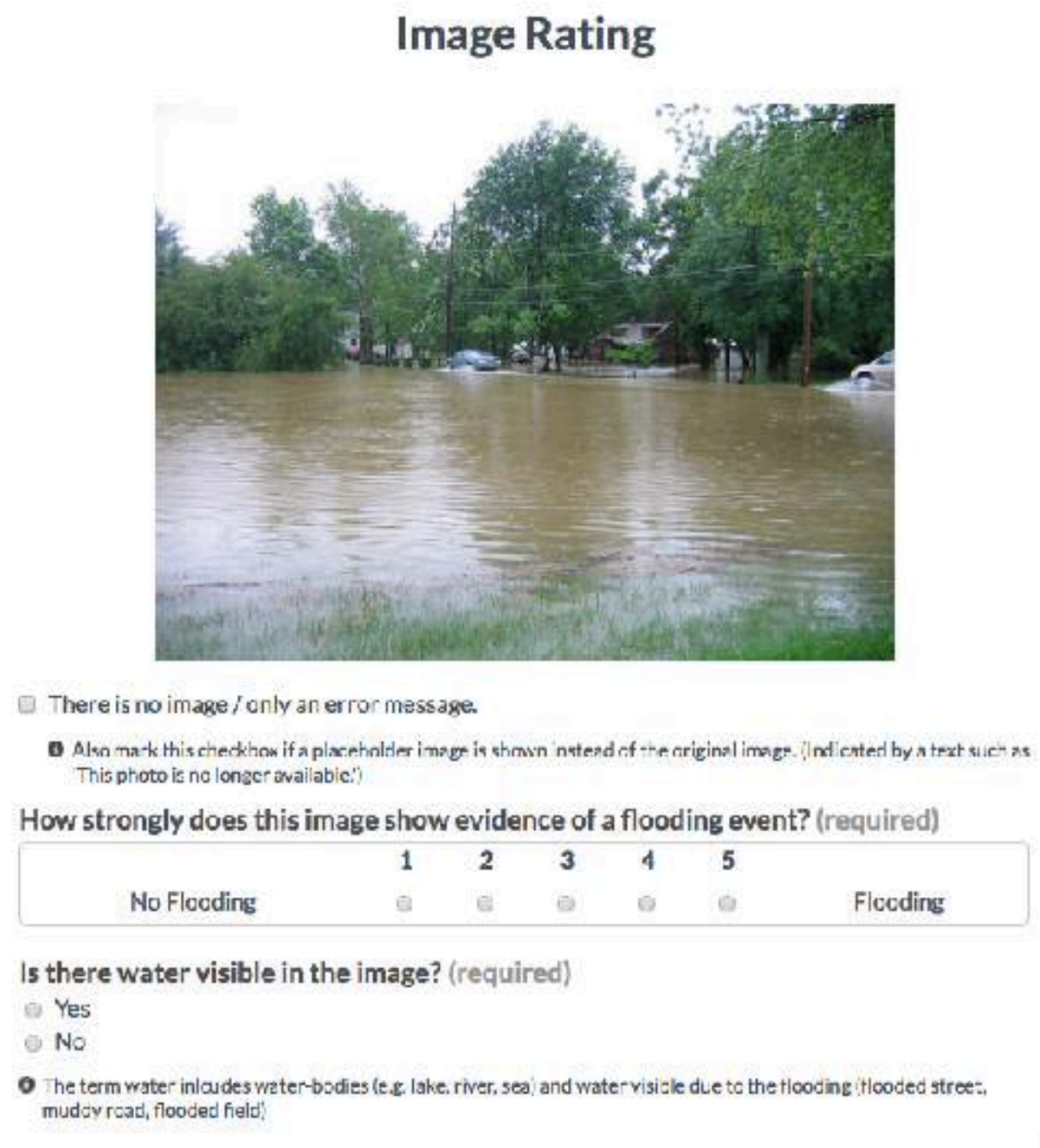
Task Dataset

- DIRSM-Dataset:
  - 6.6k images from YFCC100M + metadata (under CC-licence)
  - Basic set of precomputed features
  - Two labels (Flooding/no Flooding)

- FDSI-Dataset:
  - High resolution satellite scenes of seven flooding events provided by PlanetLabs
  - Cropped Image Patches (320x320px)
  - Segmentation masks for each patch

Ground Truth

- DIRSM:
  - Image rating according to the strength of the evidence of flooding (1-5) on Crowdflower (crowdsourcing)
  - Label as flooding if annotators rate with 4, 5 and non flooding for 1, 2
  - Additional distractor images
- FDSI:
  - Segmentation masks extracted by human annotators



Task Participation

- 15 Teams registered, 11 submitted runs:
  - from all the world (Brasil, Australia, Greece, Brunei, Italy, UK, Germany, Netherlands, Norway, Pakistan)
  - 11 teams submitted for the DIRSM subtask
  - 6 teams additionally for the FDSI subtask

- In total **63** submissions:
  - **44** submission for the first task
  - **19** submission for the second subtask

Disaster Image Retrieval From Social Media - Visual

- Retrieval only based on Image-Information

Evidence of a Flooding Event

No Evidence of a Flooding Event

DIRSM - Classification using Features from IR

SVM (RBF)	ACC	CEDD	CL	EH	FCTH	Gabor	JCD	SC	Tamara
Precision	68	75	75	**76**	72	57	74	64	65
Recall	69	75	75	**76**	73	62	74	65	66
F1-Score	68	75	75	**76**	72	54	73	60	62

- **Features:**
  - ACC - Auto-Color-Correlogram
  - CEDD - Color and Edge Directivity Descriptor
  - CL - Color Layout
  - EH - Edge Histogram
  - FCTH - Fuzzy Color and Texture Histogram
  - Gabor
  - JCD - Joint Composite Descriptor
  - SC - Scalable color
  - Tamara

- **Training**
  - 80/20 Split Training/Test
  - Baseline **63,26%**

DIRSM - Problems with traditional Features from IR

- Low-Level features are not expressive enough
  - Similar color, texture, water




DIRSM - Problems with traditional Features from IR

- Low-Level features are not expressive enough
  - Similar color, texture, water

- Can we leverage high-level features from pre-trained-CNN?
  - Fine-tuning of CNNs
  - Extracting Layers of pre-trained CNNs

DIRSM - Problems with traditional Features from IR

- Low-Level features are not expressive enough
  - Similar color, texture, water

- Can we leverage high-level features from pre-trained-CNN?
  - Fine-tuning of CNNs
  - **Extracting Layers of pre-trained CNNs**

DIRSM - Classification using Features of pre-trained CNNs

ImageNet based FeatureExtractor	lin. SVM (F1-Score)	RBF SVM (F1-Score)
AlexNet (ImageNet) - FC6	79	85
AlexNet (ImageNet) - FC7	77	83
SqueezeNet 1.0 (ImageNet) - Fire 10	**83**	**87**
SqueezeNet 1.1 (ImageNet) - Fire10	**83**	86
ResNet18 (ImageNet) - FC1000	72	70
ResNet50 (ImageNet) - FC1000	75	71
ResNet101 (ImageNet) - FC1000	77	74
ResNet150 (ImageNet) - FC1000	76	74

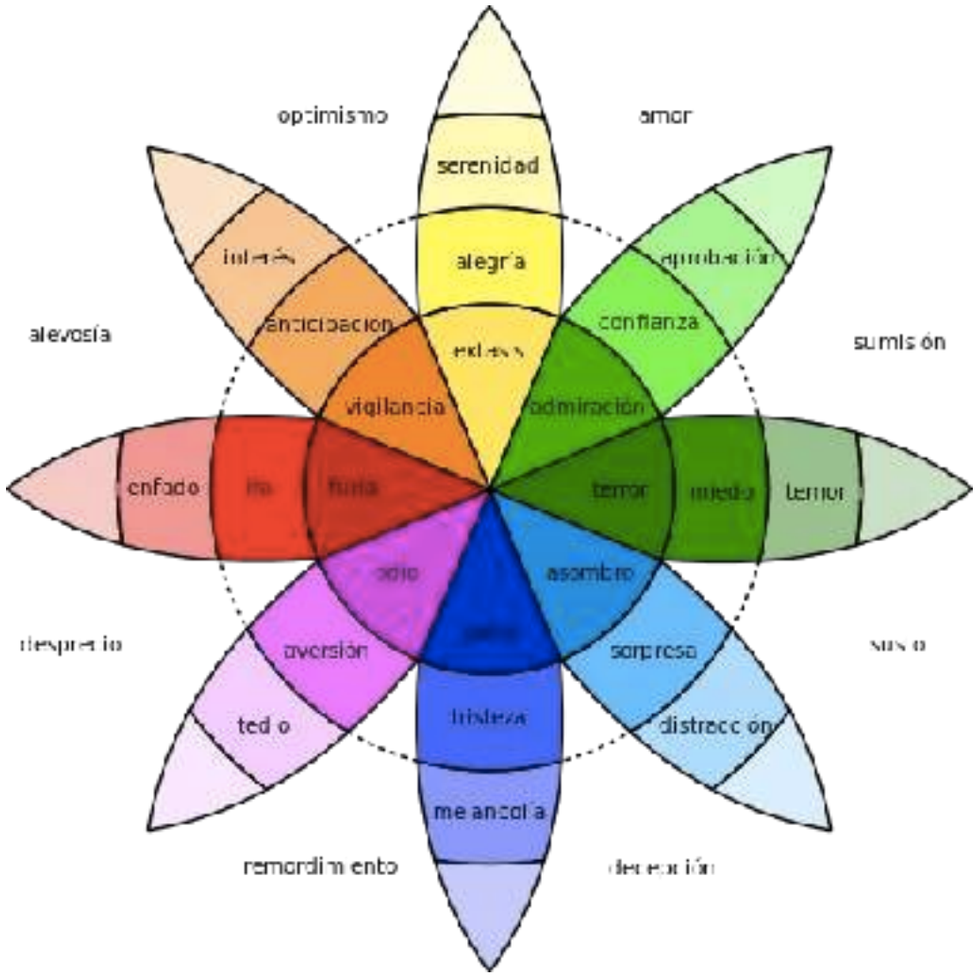
DIRSM - Classification using Features of pre-trained CNNs

ImageNet based FeatureExtractor	lin. SVM (F1-Score)	RBF SVM (F1-Score)
AlexNet (ImageNet) - FC6	79	85
AlexNet (ImageNet) - FC7	77	83
SqueezeNet 1.0 (ImageNet) - Fire 10	**83**	**87**
SqueezeNet 1.1 (ImageNet) - Fire10	**83**	86
ResNet18 (ImageNet) - FC1000	72	70
ResNet50 (ImageNet) - FC1000	75	71
ResNet101 (ImageNet) - FC1000	77	74
ResNet150 (ImageNet) - FC1000	76	74

- Are there more suitable CNN-models for this problem?

Adjective Noun Pairs with DeepSentiBank

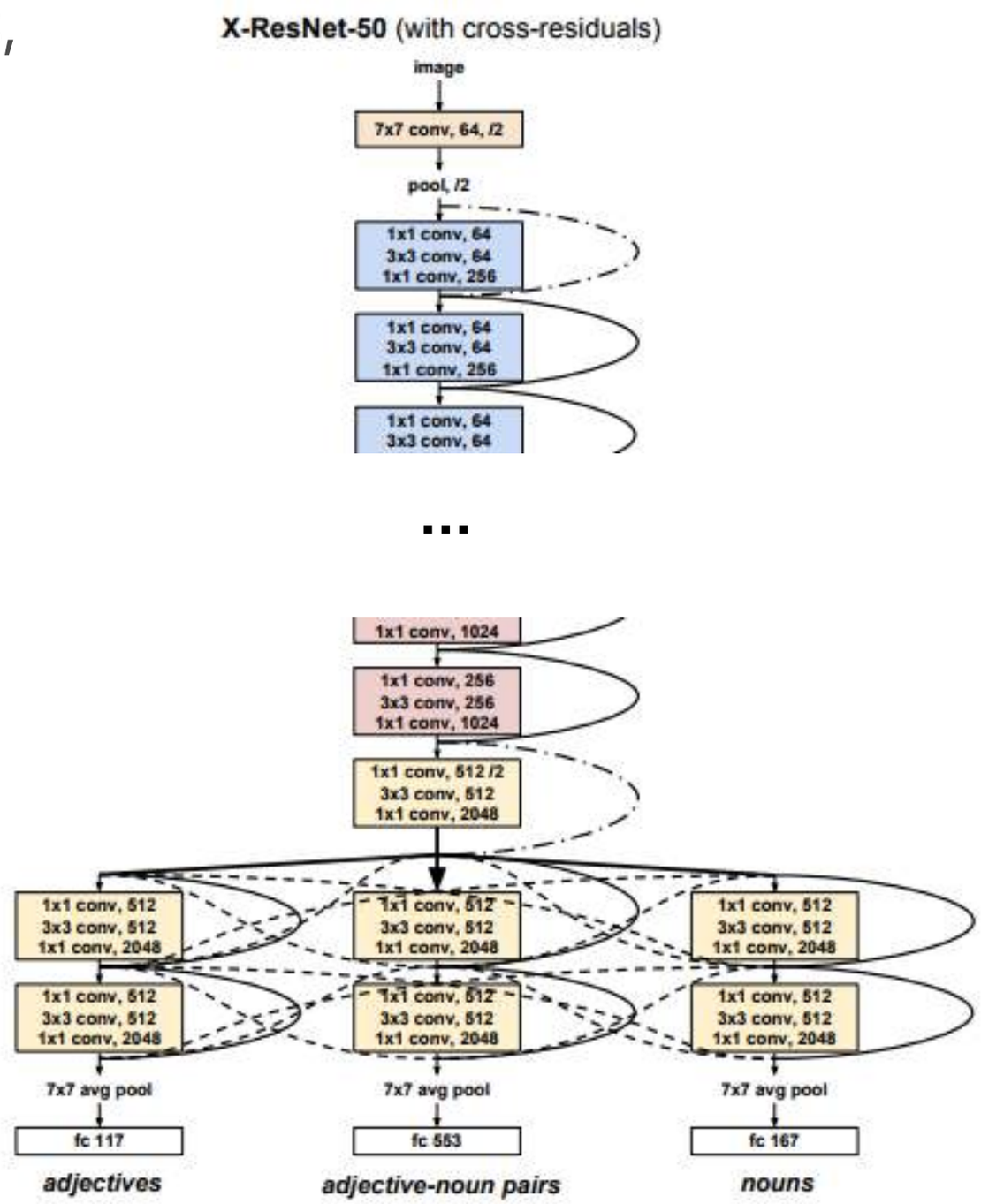
- Borth et. al. introduced Adjective Noun Pairs with the Visual Sentiment Ontology (VSO)
- DeepSentiBank uses state-of-the art CNNs to predict Adjective-Noun Pairs (ANPs)
    - AlexNet
    - VGG
    - GoogleNet
    - **X-ResNet**



- TODO: Reference

DeepSentiBank Model X-ResNet

- MultiTask prediction for Adjectives, Nouns and Adjective-Noun Pairs
- Building upon Residual Networks
- Cross-Residual Connections



- TODO: Reference

DeepSentiBank Predictions

rainy street
rainy night
wet road
wet evening

wet road
shiny sign
rainy street
damaged building

stormy clouds
dark clouds
dark landscape
small river

- Describe the scene with ANPs (*wet road, damaged building, stormy sky*)
- Less domain change (compared to ImageNet)

DIRSM - Classification using Features of pre-trained CNNs

ImageNet based FeatureExtractor	lin. SVM (F1-Score)	RBF SVM (F1-Score)
AlexNet (ImageNet) - FC6	79	85
AlexNet (ImageNet) - FC7	77	83
SqueezeNet 1.0 (ImageNet) - Fire 10	**83**	**87**
SqueezeNet 1.1 (ImageNet) - Fire10	**83**	86
ResNet18 (ImageNet) - FC1000	72	70
ResNet50 (ImageNet) - FC1000	75	71
ResNet101 (ImageNet) - FC1000	77	74
ResNet150 (ImageNet) - FC1000	76	74

VSO based FeatureExtractor	lin. SVM (F1-Score)	RBF SVM (F1-Score)
DeepSentiBank (AlexNet) - FC6	86.8	90.2
DeepSentiBank (AlexNet) - FC7	86.2	90.0
DeepSentiBank (GoogLeNet) - Pool5/7x7_s1	85.4	92.3
DeepSentiBank (XResNet) - Anptask/pool5	**90.9**	**93.7**
DeepSentiBank (XResNet) - Adjtask/pool5	89.2	**93.5**
DeepSentiBank (XResNet) - Nountask/pool5	90.2	93.2

DIRSM - Ranking using CNN-Features + SVM

- Comparing CNN-Vectors to mean class vector suboptimal
- Ranking based on distance to decision boundary of SVM with RBF-Kernel

	DSB (XResNet) - Anptask/pool5	DSB (XResNet) - Adjtask/pool5	DSB (GoogLeNet) - Pool5/7x7_s1
Average Precision at 50	**1.0**	**1.0**	0.9790
Average Precision at 100	**1.0**	**1.0**	0.9847
Average Precision at 250	0.9949	**0.9964**	0.9791
Average Precision at 500	0.9869	**0.9875**	0.9683
Average Precision at 1156	**0.9547**	0.9538	0.9239
Average Precision at mean K	0.9873	**0,9875**	0,9687

- Only 40% (2111) training, 60% (3168) test

Disaster Image Retrieval

- Sample Predictions & Results

Disaster Image Retrieval From Social Media - Metadata

- Retrieval only based on Meta-Information


```
{ "images":[{ "image_id": "12328463323",
              "image_url": "http://www.flickr.com/photos/9752474@N07/12328463323/",
              "image_extension_original": "jpg",
              "date_taken": "2014-01-30 10:18:12.0",
              "date_uploaded": "1391631137",
              "user_nsid": "9752474@N07",
              "user_nickname": "SurferJoe88",
              "title": "Emma Wood State Beach - campsites",
              "description": "Emma Wood State Beach - flooded campsites",
              "user_tags": ["flooding"],
              "license_name": "Attribution-NonCommercial-ShareAlike License",
              "license_url": "http://creativecommons.org/licenses/by-nc-sa/2.0/",
              "capture_device": "SAMSUNG PL70 / VLUU PL70 / SAMSUNG SL720",
              "latitude": 34.28715400000000102,
              "longitude": -119.32989299999999844},
              ...
            ]
}
```

Disaster Image Retrieval - Metadata

- Evaluations of Title, Description, UserTags
- Word embeddings with word2Vec (200 dimensions) on UserTags
- Document representation as mean word embeddings
- Importance Weighting of Word Embeddings with term frequency document inverse frequency (TF-IDF)
- 63.41% AP on Testset

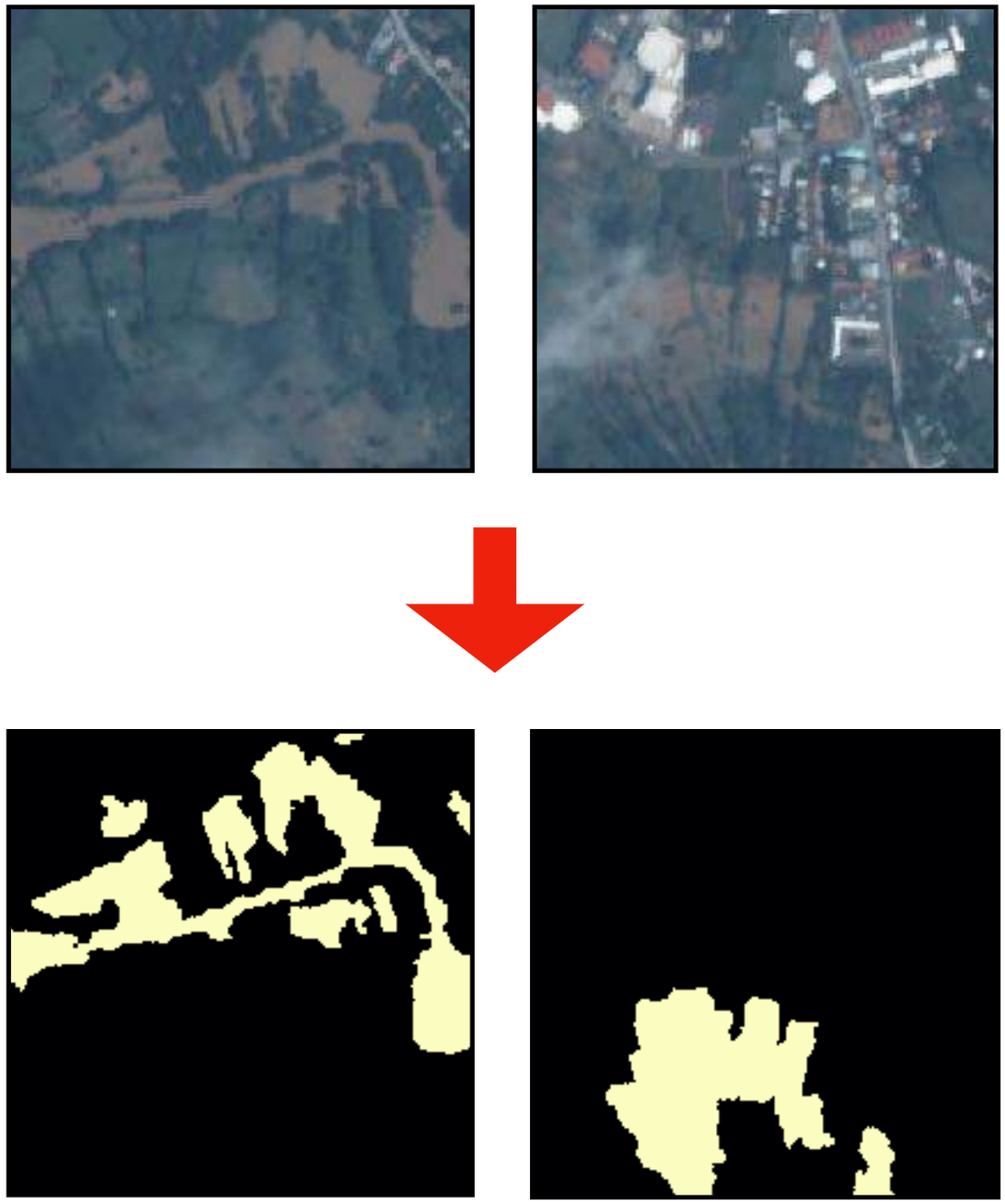
DIRSM - Visual and Metadata Information

- Fusion of Embeddings and Visual Info on Feature-level
- Concatenation of Image Feature and Word Embedding

	AP@480	MAP@[50, 100, 150, 240, 480]
Visual (XResNet)	86.64	95.71
Metadata	63.41	77.64
Visual & Metadata	**90.45**	**97.40**
Visual (ImageNet)	74.08	64.50

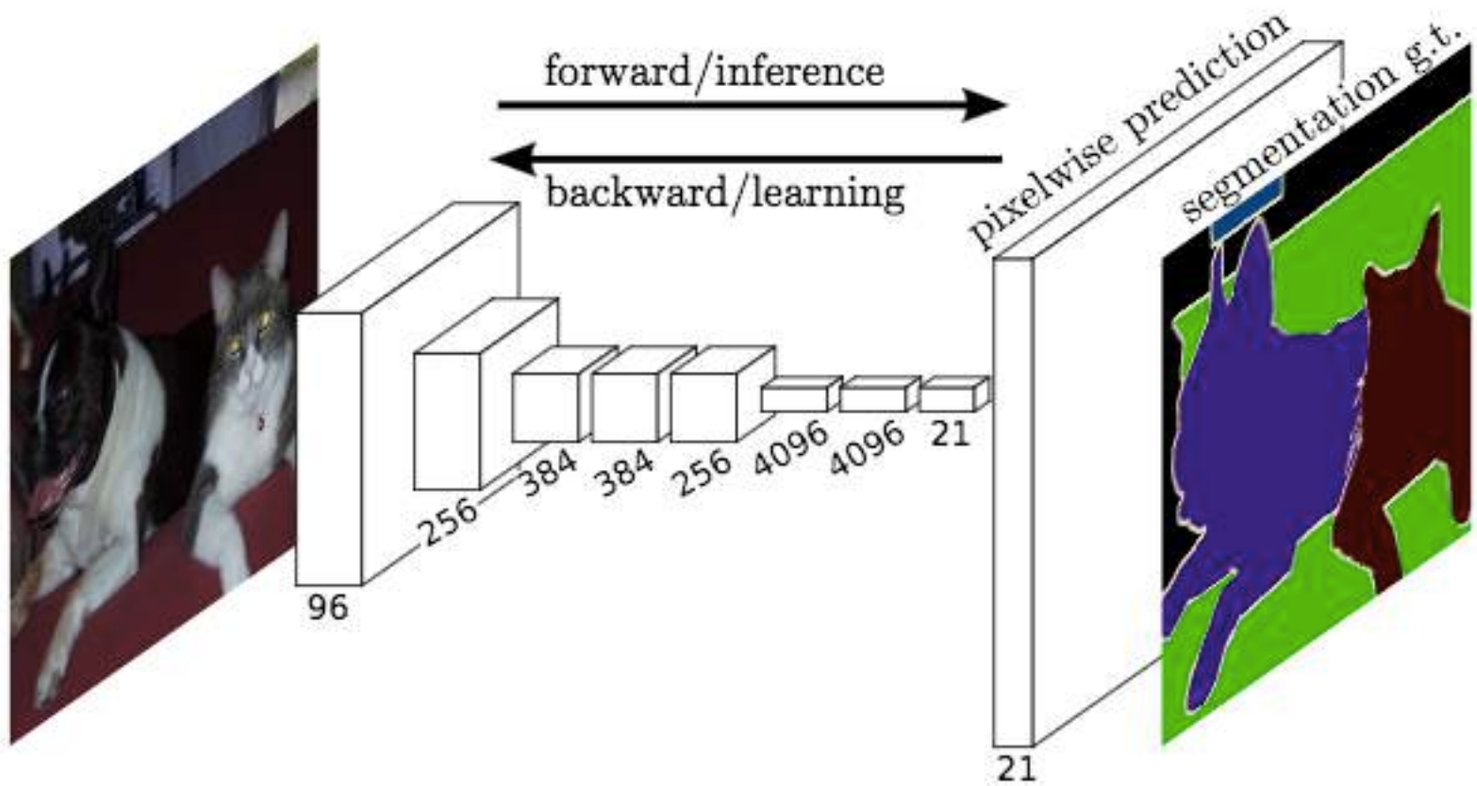
Flood Detection from Satellite Images

- Many approaches extract pixel information from Remote Sensing and traditional classifiers
- Can we apply networks used for Semantic Segmentation?
- Can we incorporate additional channels?



From Classification to Semantic Segmentation

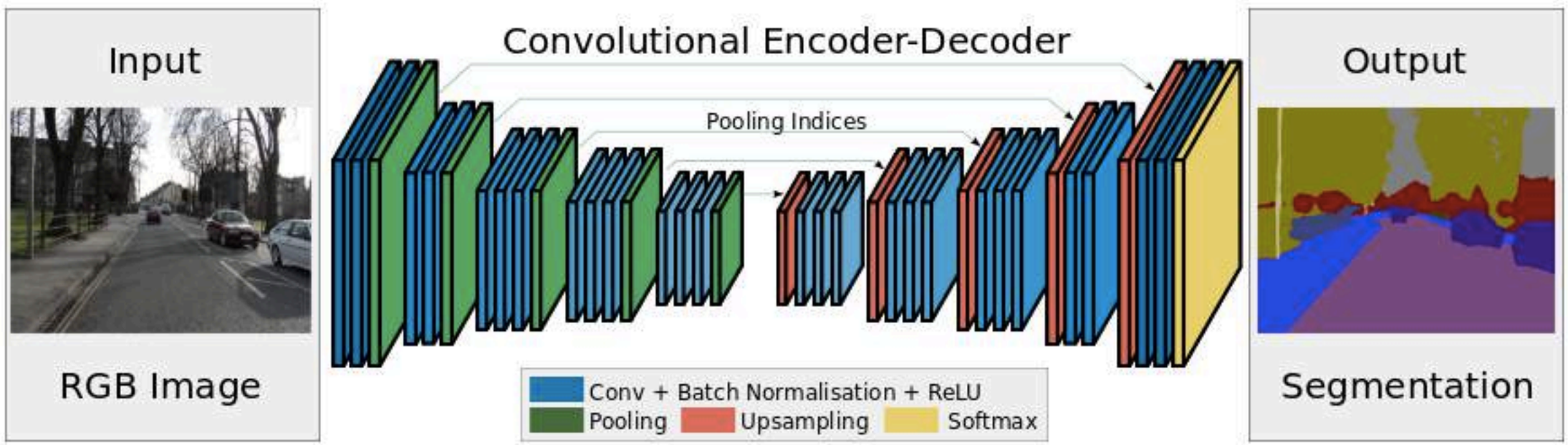
- Fully Convolutional Networks
- Pixel-wise predictions by changing fully connected to fully convolutional layers



- TODO: Reference

Network Architectures for Semantic Segmentation

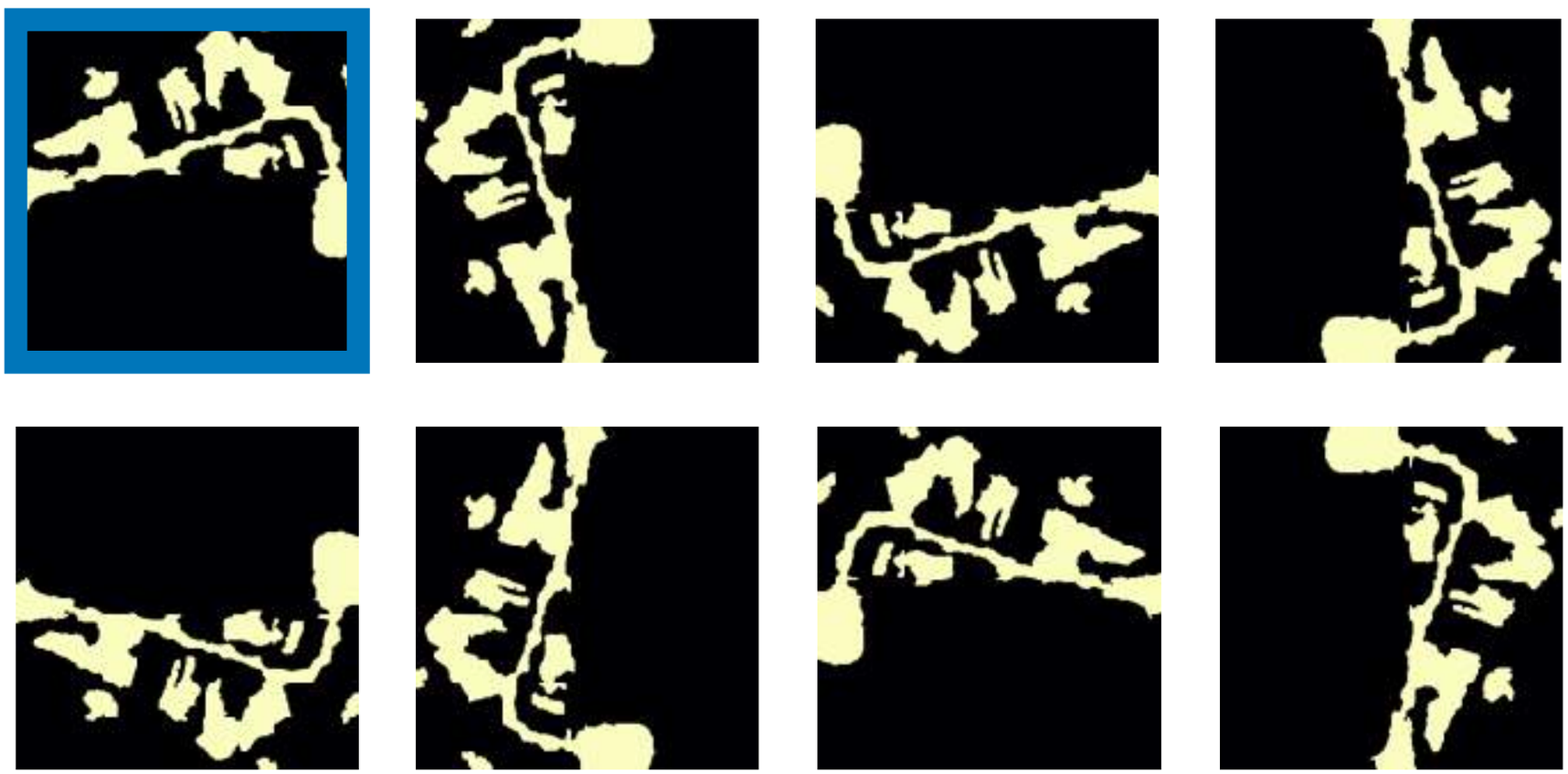
- Fully Convolutional Network (RGB)
- Fully Convolutional Network (RGB + IR)
- SegNet (RGB + IR)



- TODO: Reference

FDSI - Data Processing

- 90 Deg. rotations with vertical and horizontal flipping
- Normalisation based on location
  - mean per channel
  - conversion of 16 bit (geoTiffs) to 8 bit



FDSI - Results

- Evaluation Results on the Testset
  - Importance of the Infrared Channel for Flood Detection
  - Importance of the Decoder

	Same Locations	New Location
FCN (RGB)	73.56	69.32
FCN (RGB+IR)	84.27	70.87
SegNet (RGB+IR)	**84.36**	**74.13**

From Flood Detection to Emergency Response with Satellite Data

From Flood-Detection to Emergency Response

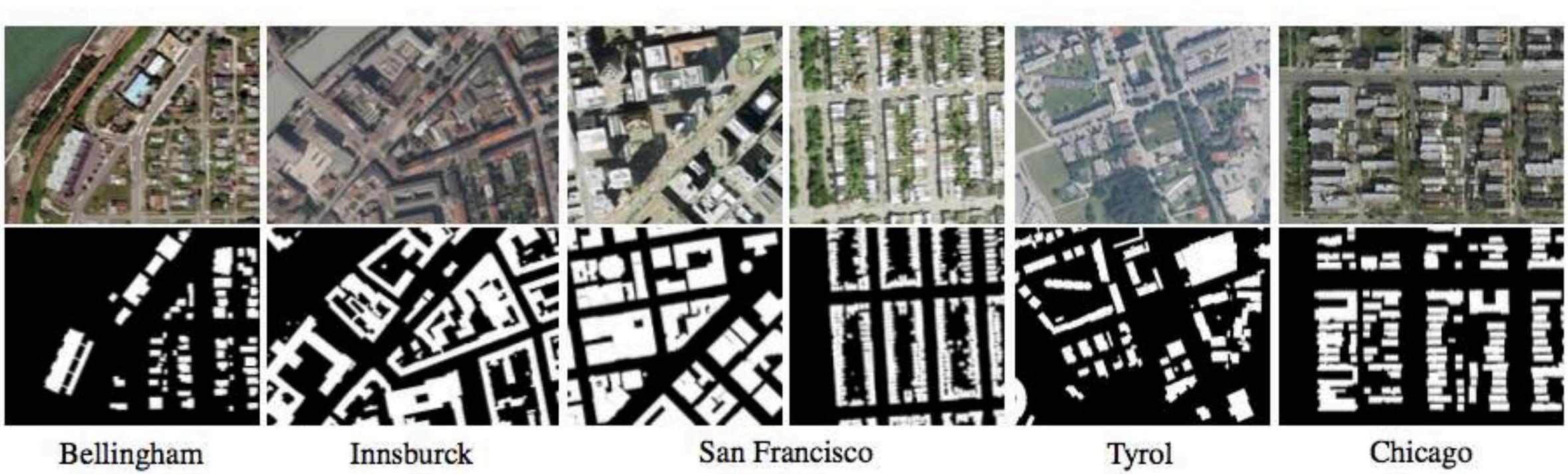
From Flood-Detection to Emergency Response

Problems with many Build Datasets

- Many publications on Building Segmentation from Satellite Data
  - Low generalisation to new places and cities
  - Many Variations (Building Types, Atmospheric Correction, Density)

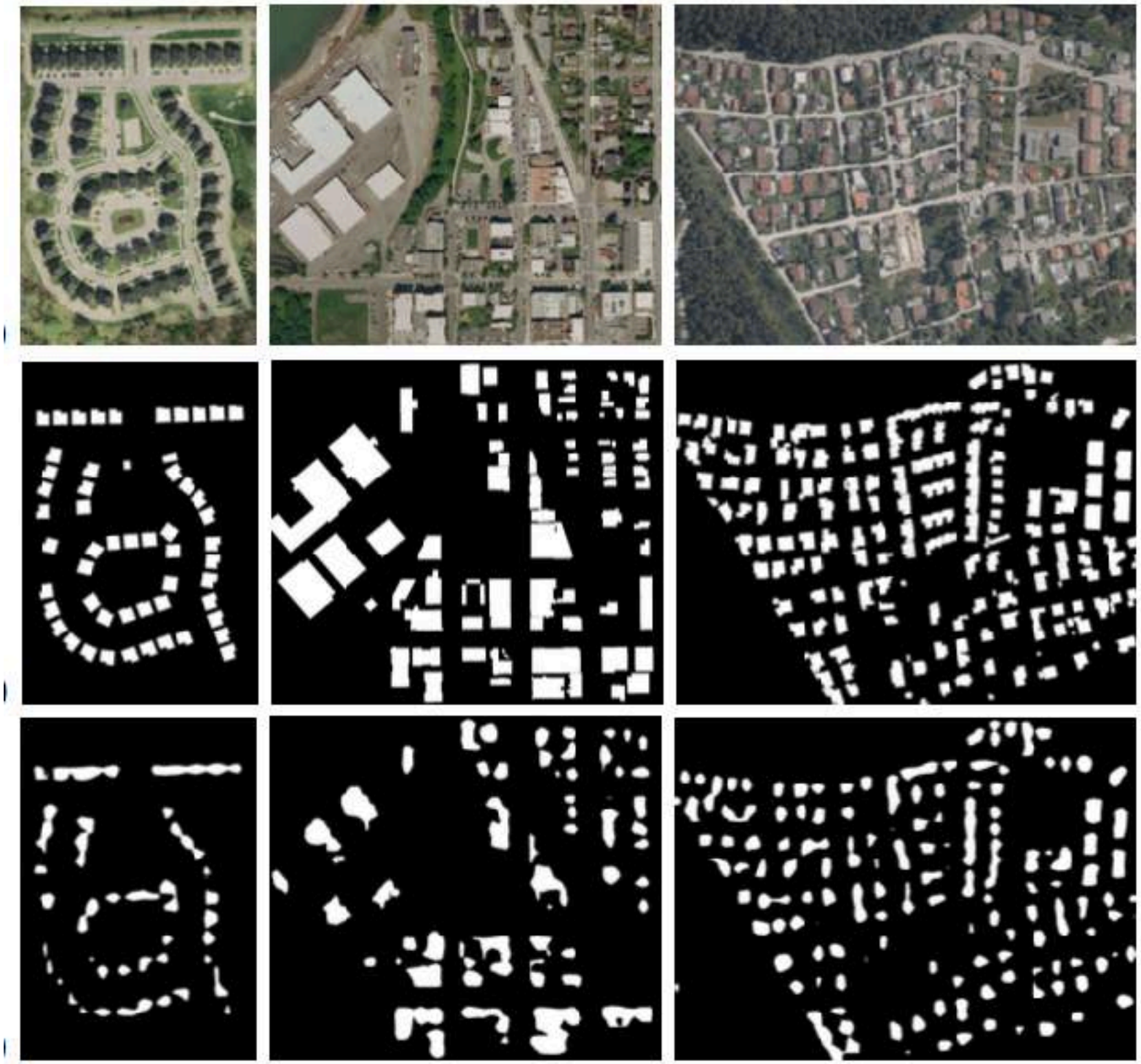
Segmentation of Building Footprints - Overview

- Large-Scale Inria Aerial Image Labeling Dataset released in May 2017 with High Resolution Imagery
  - Specifically addressing Variations
- Open Challenge with Leaderboard
- Evaluated different CNN-architectures for Sem. Segmentation

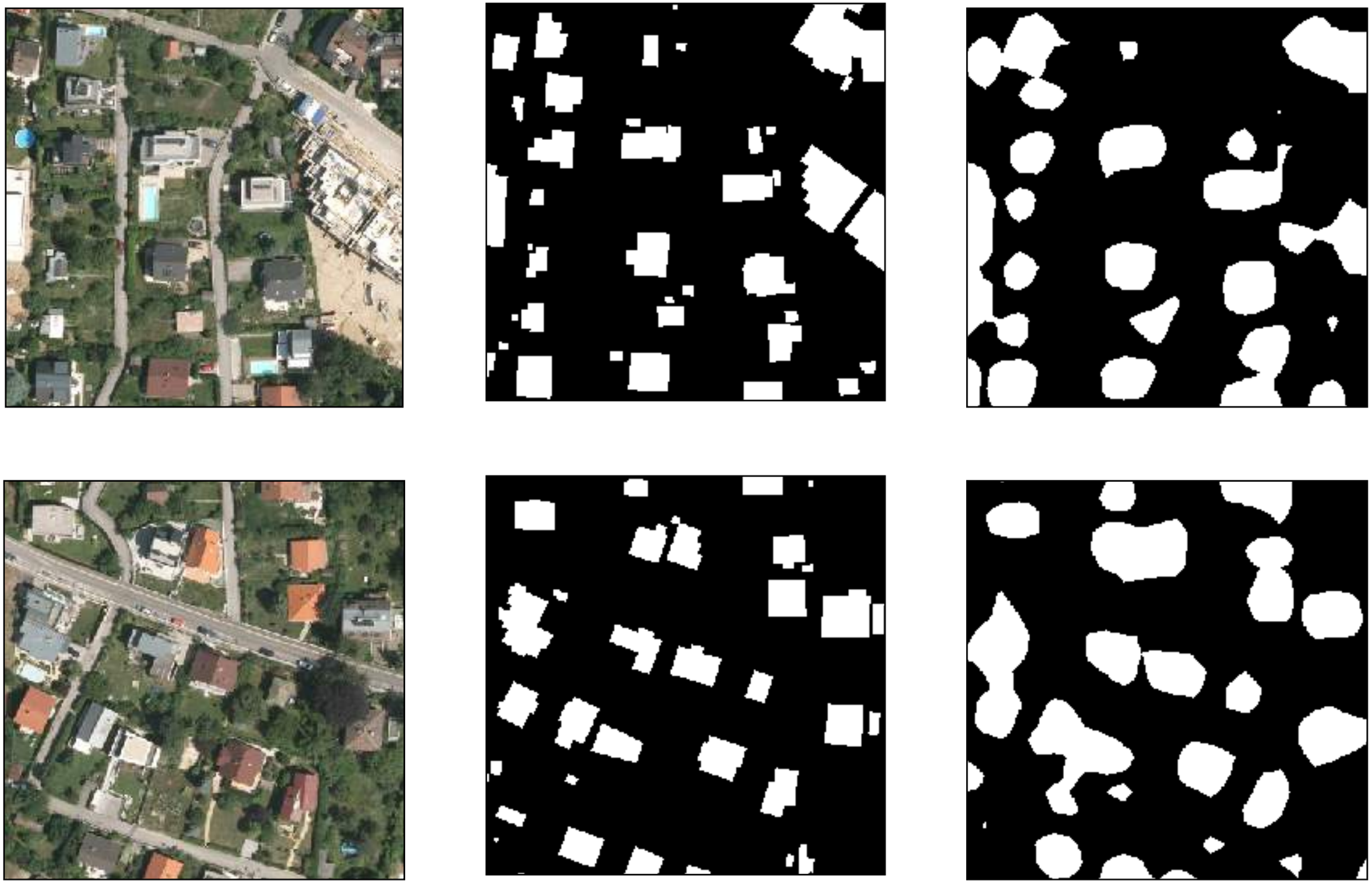


- TODO: Reference

Segmentation of Building Footprints - „Blobby“-Predictions!

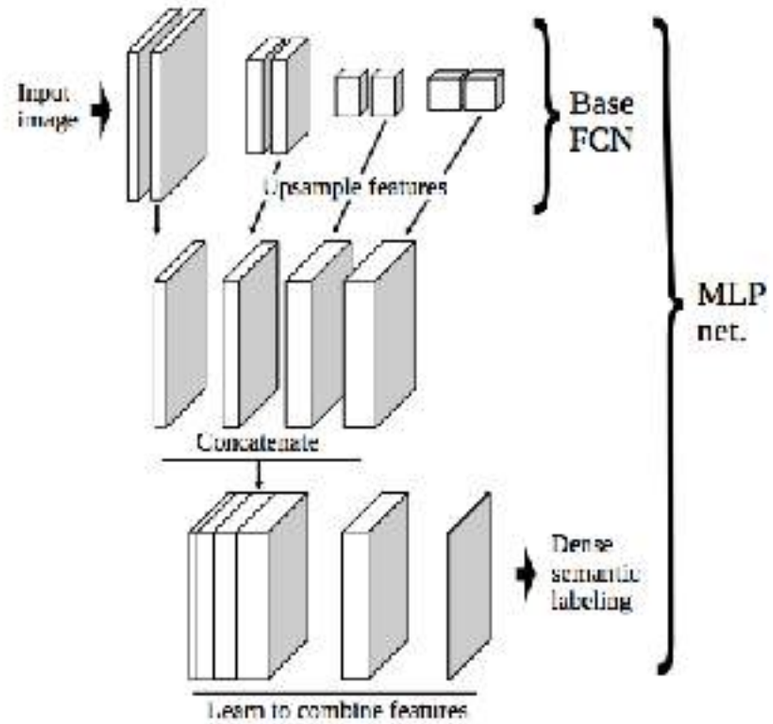


Segmentation of Building Footprints - „Blobby“-Predictions!



Segmentation of Building Footprints - Baseline Architectures

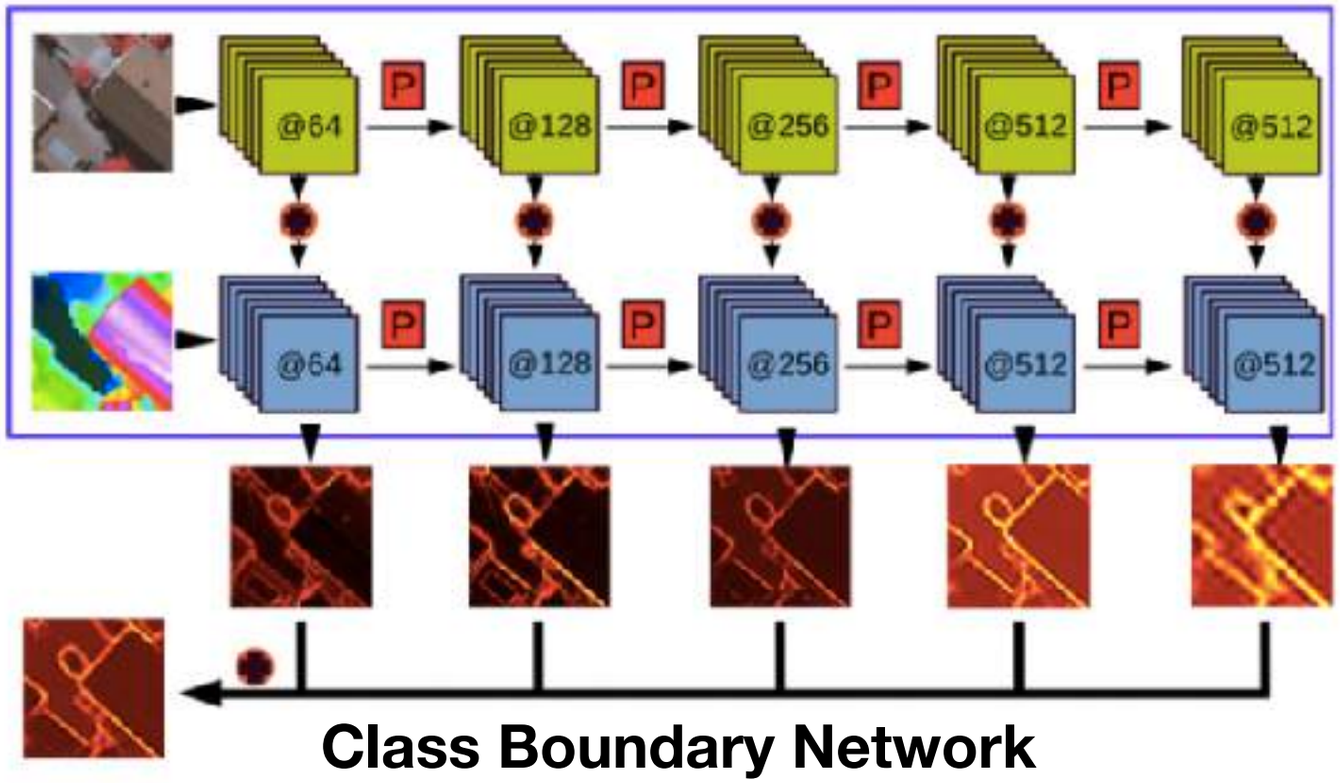
- Different Architectures
  - FCN
  - Unpooling-Net (SegNet)
  - Skip-Net (U-Net)
  - FCN + MLP



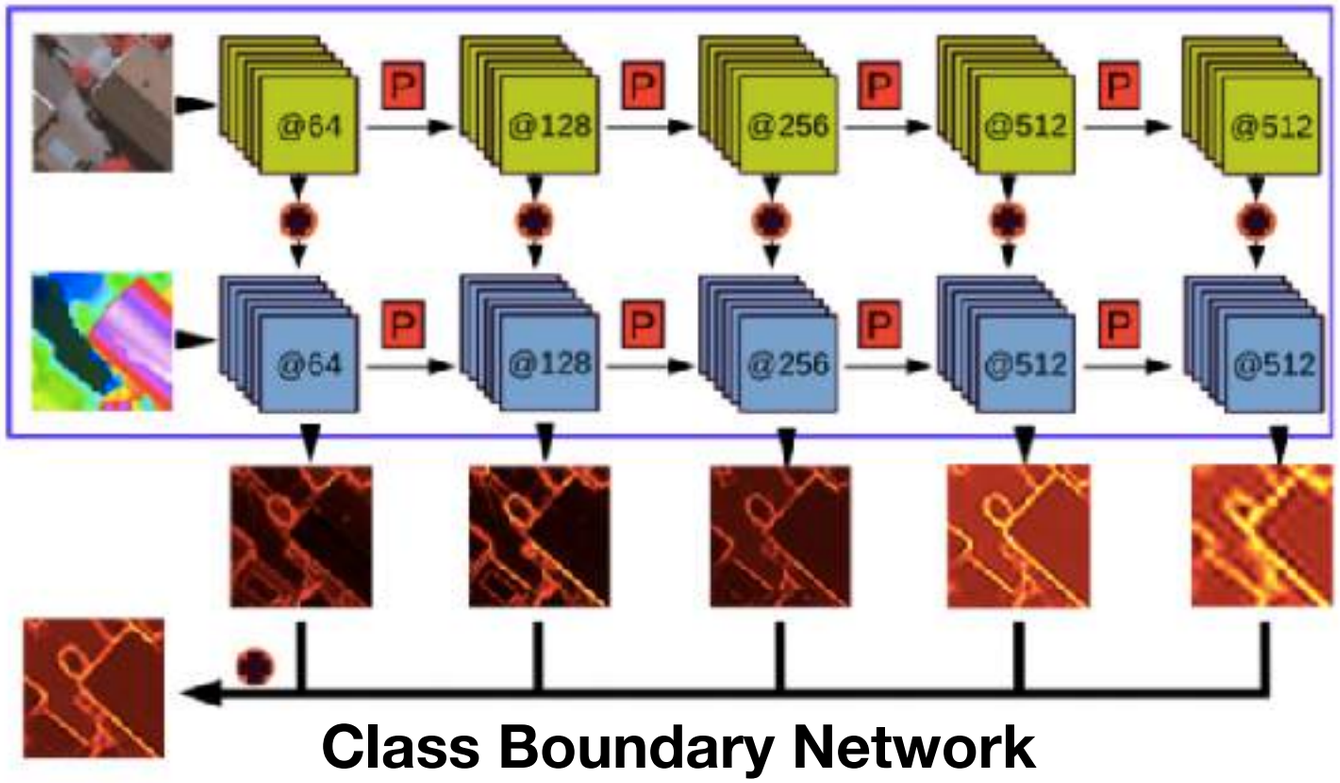
- Results

		Austin	Chicago	Kitsap Co.	West Tyrol	Vienna	Overall
FCN	IoU	47.66	53.62	33.70	46.86	60.60	53.82
	Acc.	92.22	88.59	98.58	95.83	88.72	92.79
Skip	IoU	57.87	61.13	46.43	54.91	70.51	62.97
	Acc.	93.85	90.54	98.84	96.47	91.48	94.24
MLP	IoU	61.20	61.30	51.50	57.95	72.13	64.67
	Acc.	94.20	90.43	98.92	96.66	91.87	94.42

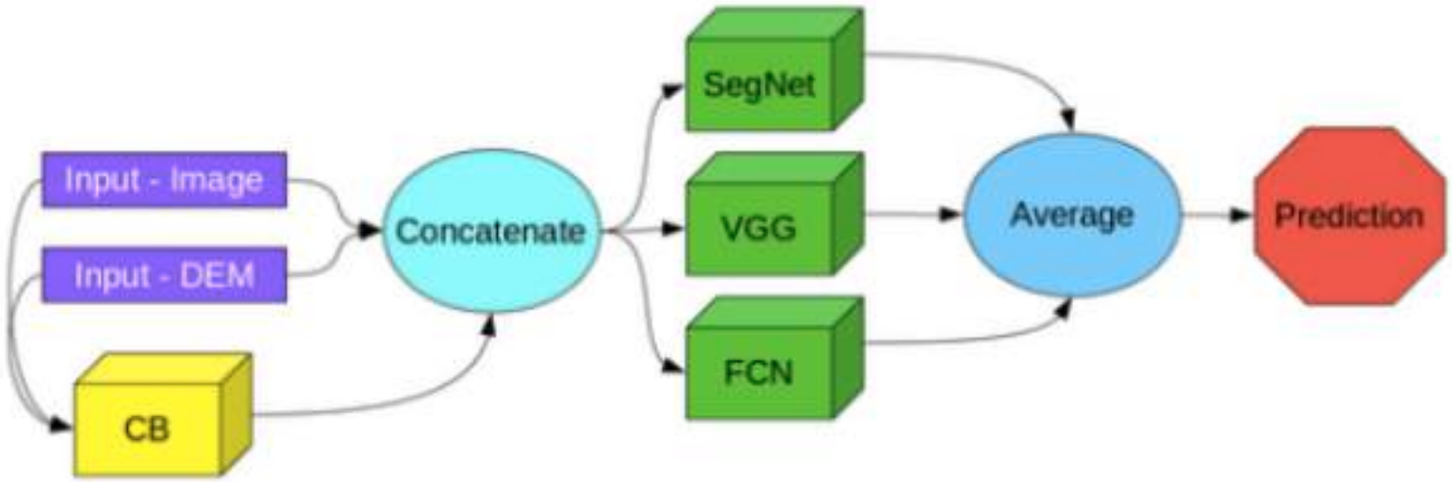
Using Multiple Networks for a Better Prediction



Using Multiple Networks for a Better Prediction



Class Boundary Network

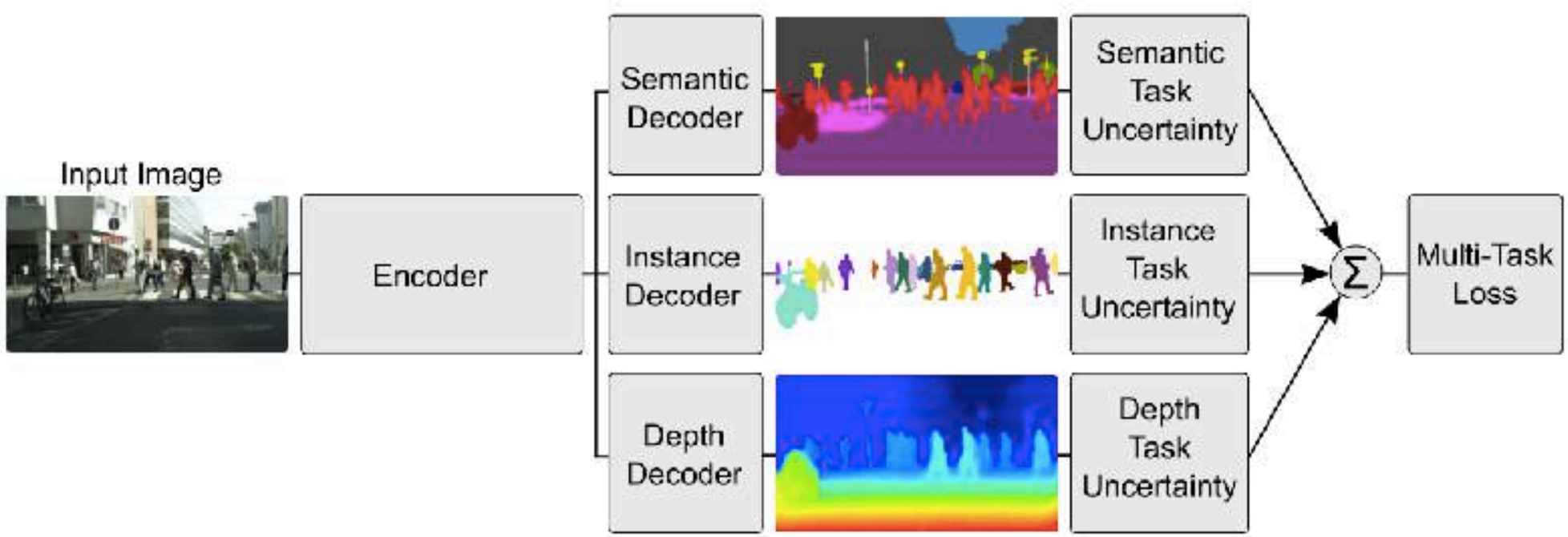


Model Ensemble of multiple Networks

- TODO: Reference

Multi-Task Learning

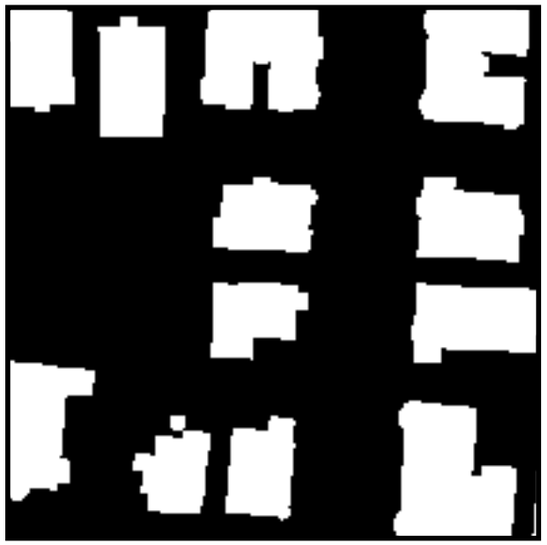
- Use complementary tasks to improve the original task
  - Implicit Data Augmentation
  - Attention Focusing
  - Representation Bias
  - Regularization



- TODO: Reference

Standard Output Representations

Input



Semantic Information

Additional Output Representations

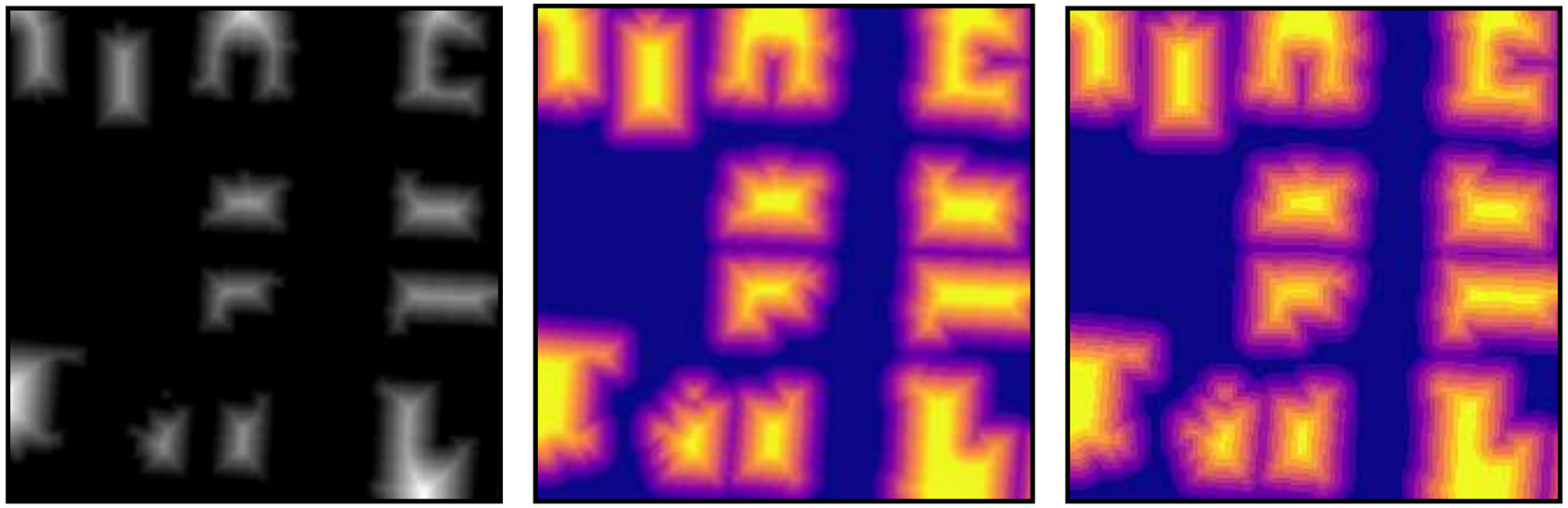
Input

$$D(p) = \delta_p min(\min_{\forall q \in Q} d(p,q), R),$$

$$\delta_p = \begin{cases} +1 & \text{if} \quad p \in C_{building} \\ -1 & \text{if} \quad p \notin C_{building} \end{cases}$$

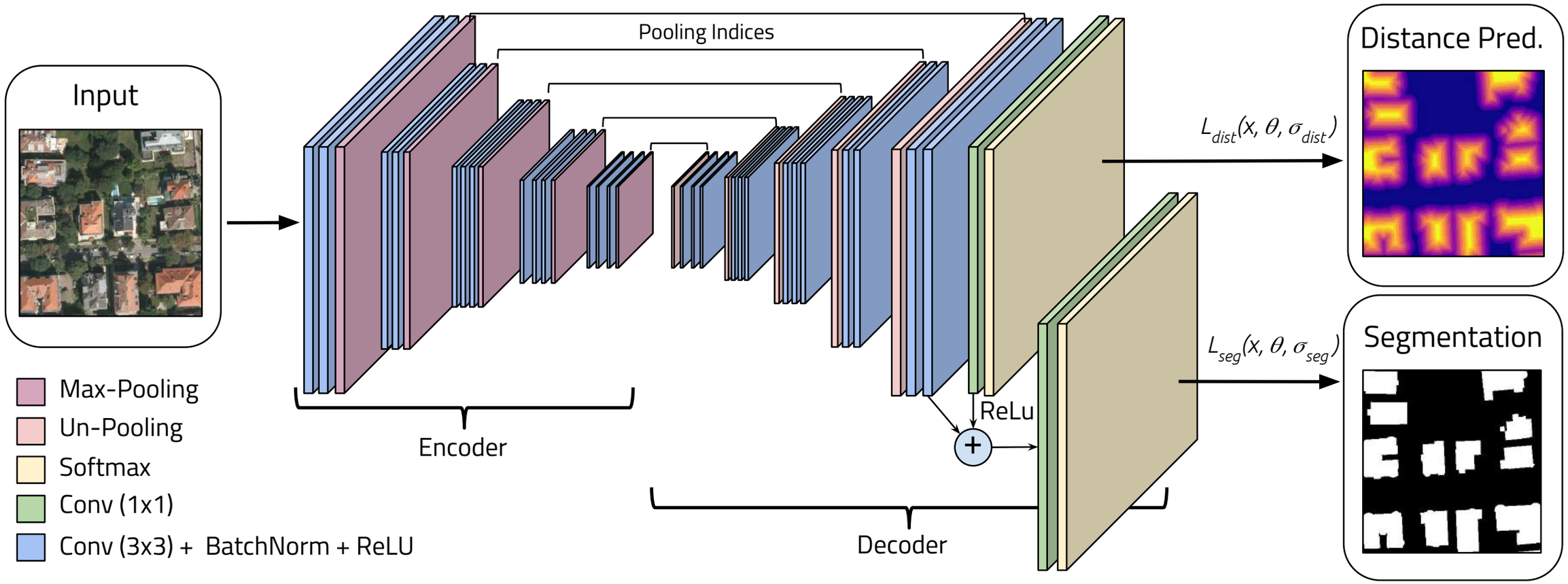
$$D(p) \sum_{k=1}^{K} r_n b_k(p) \sum_{k=1}^{K} b_k(p) = 1$$

Semantic Information



Geometric Information (Distance)

Cascaded Multi-Task Network



$$L_{total}(x; \theta, \sigma_{dist}, \sigma_{seg}) = L_{dist}(x; \theta, \sigma_{dist}) + L_{seg}(x; \theta, \sigma_{seg})$$

Uncertainty Based Task Weighting

- Model Task weights as Uncertainty of each Task
  - Optimize network parameters and uncertainty parameters together

$$L_t(x, \theta, \sigma_t) = \sum_{c=1}^{C} -C_c log P(C_c = 1 | x, \theta, \sigma_t)$$

$$= \sum_{c=1}^{C} -C_c log(exp(\frac{1}{\sigma_t^2} f_c(x))) + log \sum_{c'=1}^{C} exp(\frac{1}{\sigma_t^2} f_{c'}(x)) \quad (6)$$

Applying the same assumption as in [25]:

$$\frac{1}{\sigma^2} \sum_{c'} exp(\frac{1}{\sigma^2} f_{c'}(x)) \approx (\sum_{c'} exp(f_{c'}(x)))^{\frac{1}{\sigma^2}} \quad (7)$$

allows to simplify Eq. 6 to:

$$L_t(x, \theta, \sigma_t) \approx \frac{1}{\sigma_t^2} \sum_{c=1}^{C} -C_c log P(C_c = 1 | x, \theta) + log(\sigma_t^2) \quad (8)$$

Quantitative Results

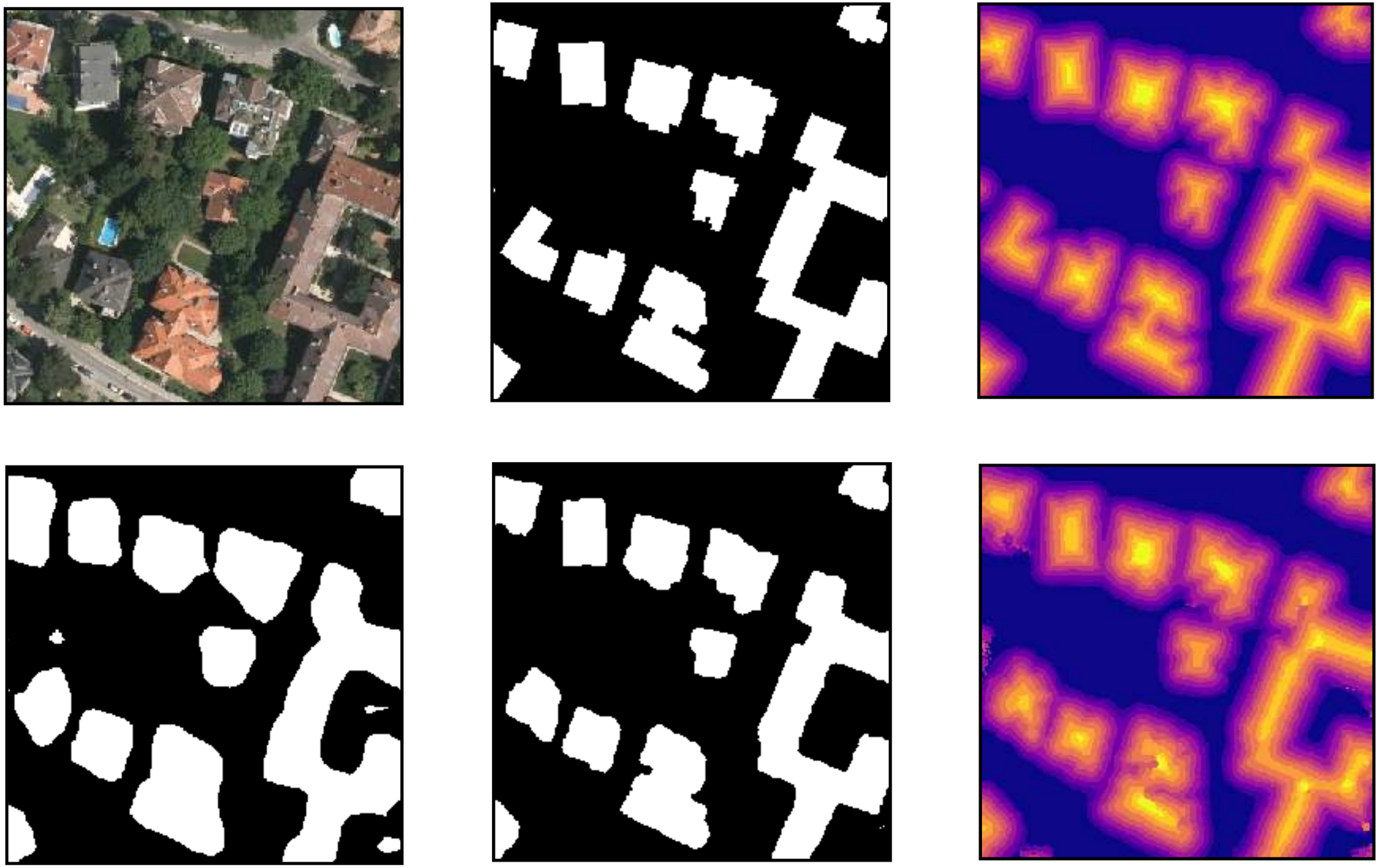
	mean IoU	Acc. (Pixel)
Baseline FCN [2]	53.82%	92.79 %
Baseline FCN + MLP[2]	64.67%	94.42 %
FCN (VGG16 encoder)	66.21%	94.54 %
FCN + MLP (VGG16 encoder)	68.17%	94.95 %
SegNet (VGG16 encoder)	**70.14%**	**95.17** %

Quantitative Results

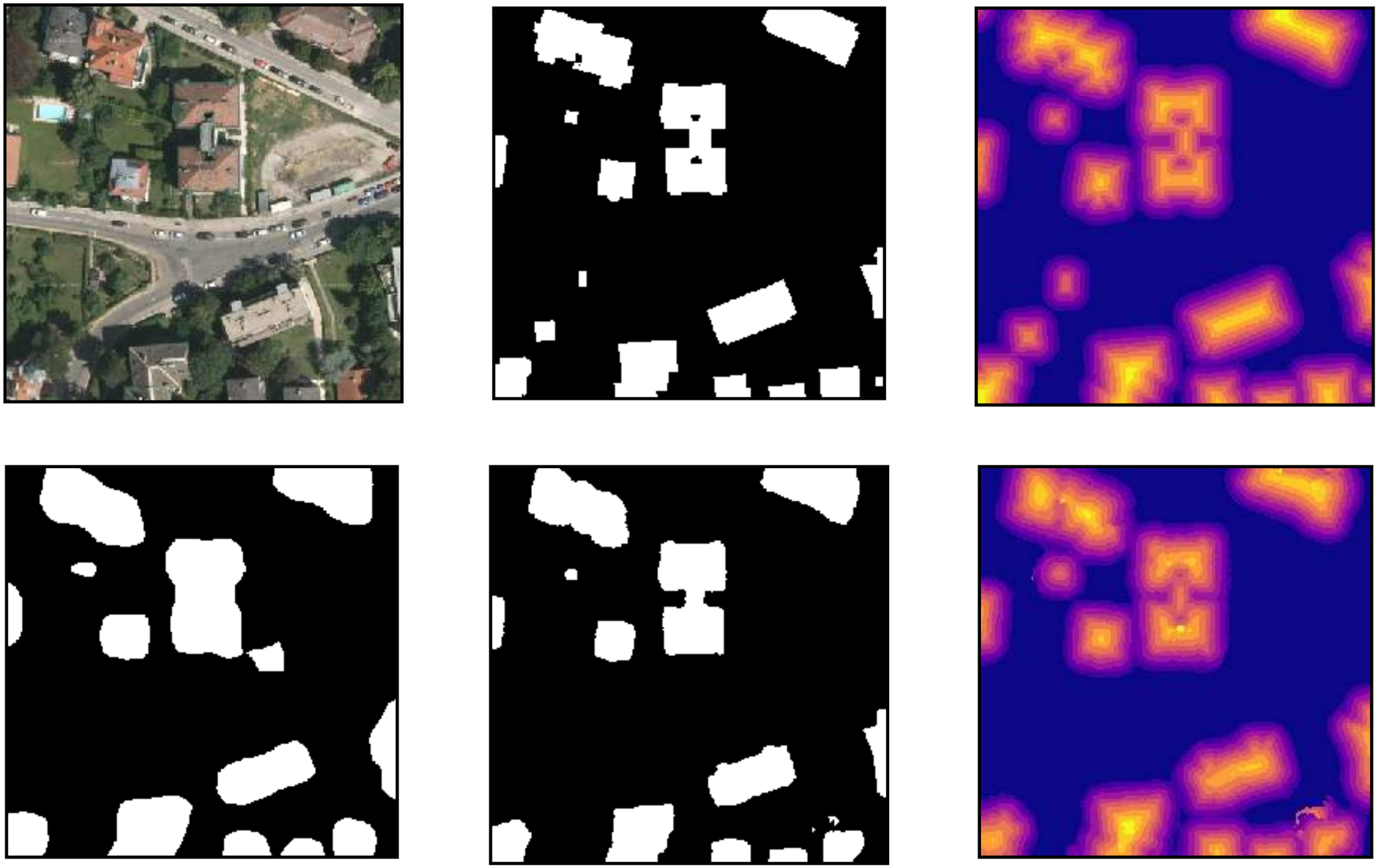
	mean IoU	Acc. (Pixel)
Baseline FCN [2]	53.82%	92.79 %
Baseline FCN + MLP[2]	64.67%	94.42 %
FCN (VGG16 encoder)	66.21%	94.54 %
FCN + MLP (VGG16 encoder)	68.17%	94.95 %
SegNet (VGG16 encoder)	**70.14%**	**95.17** %

		Austin	Chicago	Kitsap Co.	West Tyrol	Vienna	Overall
FCN + MLP	IoU	61.20	61.30	51.50	57.95	72.13	64.67
(Baseline)	Acc.	94.20	90.43	98.92	96.66	91.87	94.42
SegNet (Single-Loss)	IoU	74.81	52.83	68.06	65.68	72.90	70.14
NLL-Loss for Seg. Classes	Acc.	92.52	98.65	97.28	91.36	96.04	95.17
SegNet (Single-Loss)	IoU	76.49	66.77	72.69	66.35	76.25	72.57
NLL-Loss for Dist. Classes	Acc.	93.12	99.24	97.79	91.58	96.55	95.66
SegNet + MultiTask-Loss	IoU	76.22	66.64	71.70	**67.03**	**76.68**	72.65
(Equally Weighted)	Acc.	93.03	99.24	97.71	91.66	96.60	95.65
SegNet + MultiTask-Loss	IoU	**76.76**	**67.06**	**73.30**	66.91	**76.68**	**73.00**
(Uncertainty Weighted)	Acc.	**93.21**	**99.25**	**97.84**	**91.71**	**96.61**	**95.73**

Qualitative Results



Qualitative Results



Acknowledgments

NVIDIA

planet.