Jet Propulsion Laboratory
California Institute of Technology

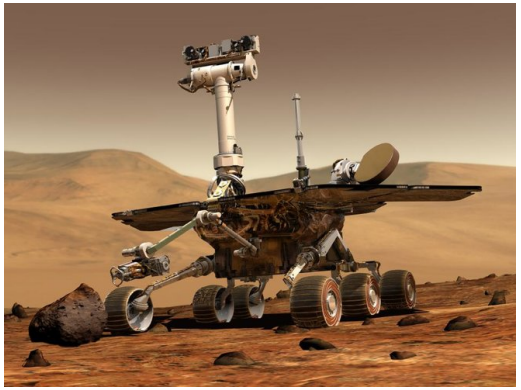
TU-Net and TDeepLab: Deep Learning-based Terrain Classification Robust to Illumination Changes, Combining Visible and Thermal Imagery

Yumi Iwashita, Adrian Stoica (NASA Jet Propulsion Laboratory, USA)

Kazuto Nakashima, Ryo Kurazume (Kyushu University, Japan)

Terrain classification (TC)

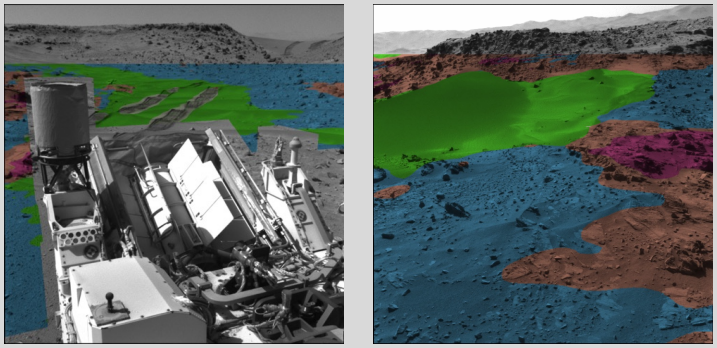
- Terrain classification (TC) is an essential component of a broader understanding of the terrain to be traversed by rovers



Spirit got stuck in a sand trap, 2010

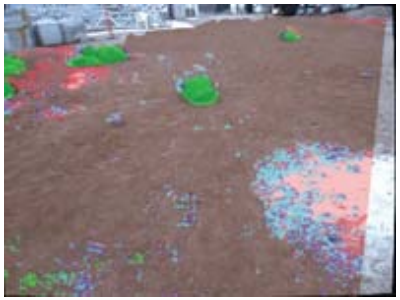
Wheels of Curiosity rover got damaged

Terrain classification from **visible** images



Sand Smooth Outcrop
Rocks Rocks on outcrop

- We have developed a TC system from visible images



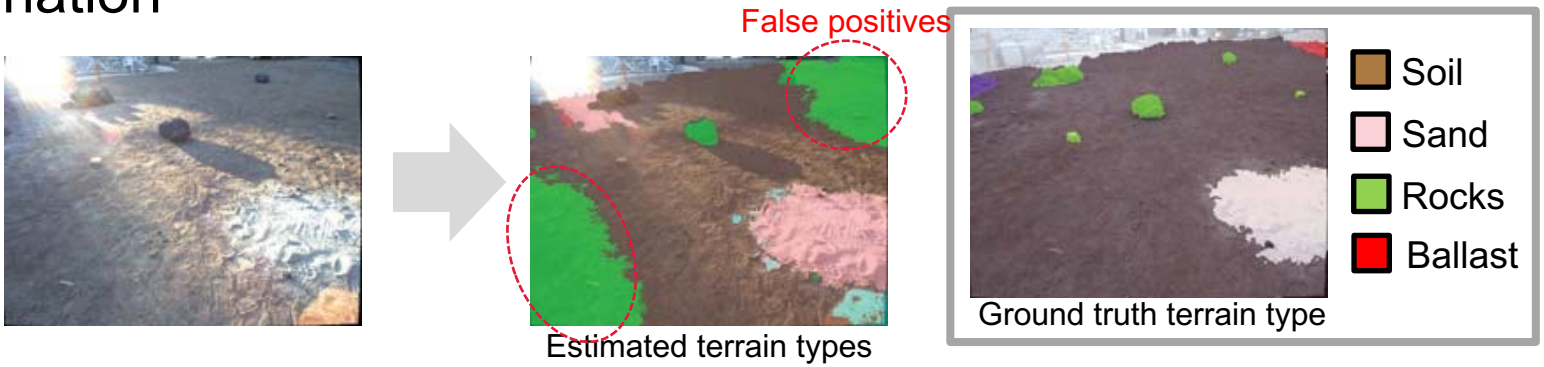
Estimated terrain type

Soil Rocks Rocky terrain
Ballast Bedrock Sand

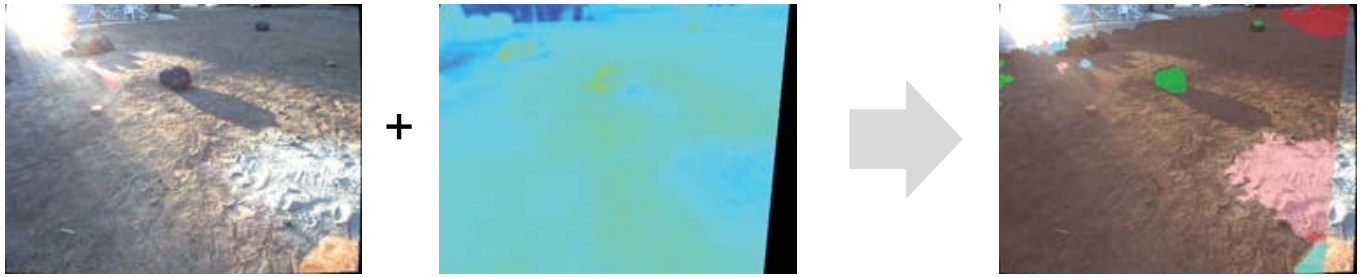
Ground truth

Terrain classification

- Terrain classification with visible information is challenging due to a variety of factors, of most importance being the changes in illumination



- The terrain classification can be aided by the fact that the temperature of various types of terrains depends on the thermal characteristics of the terrain



Captured RGB image | Captured IR image | Estimated terrain types

Deep learning for terrain classification

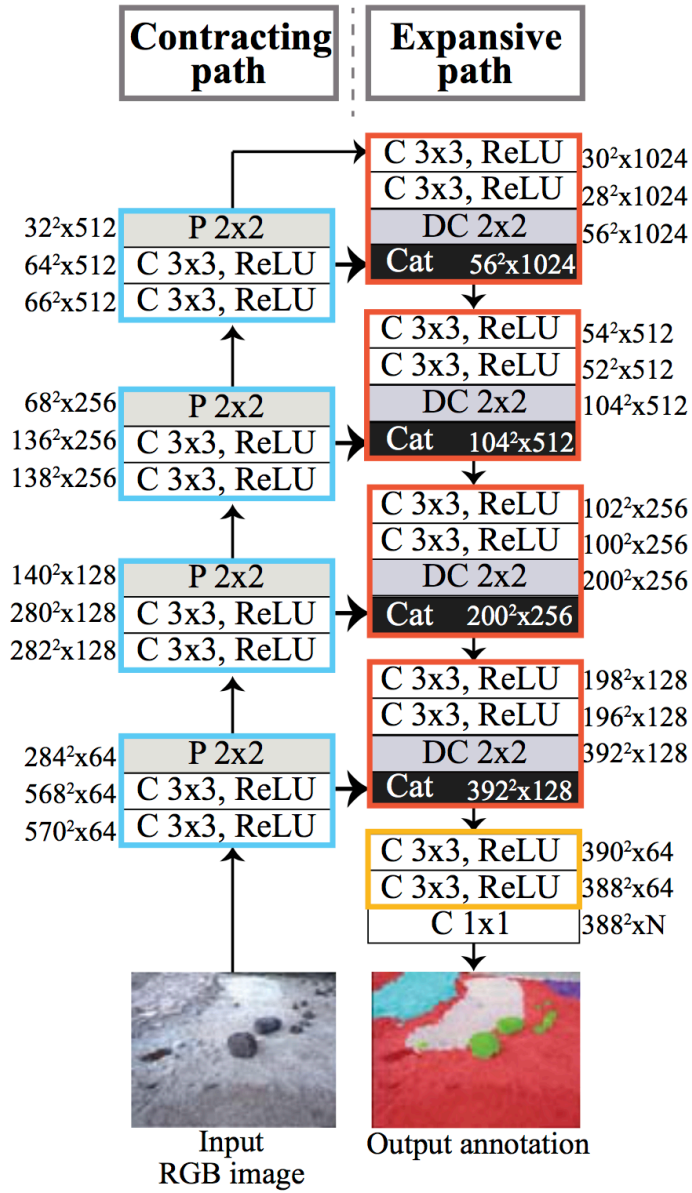
- Existing approaches
  1. Fully Convolutional Networks (FCN), such as DeepLab
     - In general it requires a huge dataset to train its parameters
       - If we do not have enough dataset to train FCN from scratch, finetune the parameters from pre-trained parameters
  2. U-Net
     - It has short-cut connections, which help a lot with parameter training even given a small number of dataset

- We propose two novel deep learning-based terrain classification methods based on U-Net and FCN (DeepLab), which combine visible and thermal IR images
  1. **TU-Net (Two U-Net)**
  2. **TDeepLab (Two Deeplab)**
- We also extend the proposed methods to Siamese-based approaches to realize robustness to illumination changes

U-Net for terrain classification from RGB images

U-Net architecture

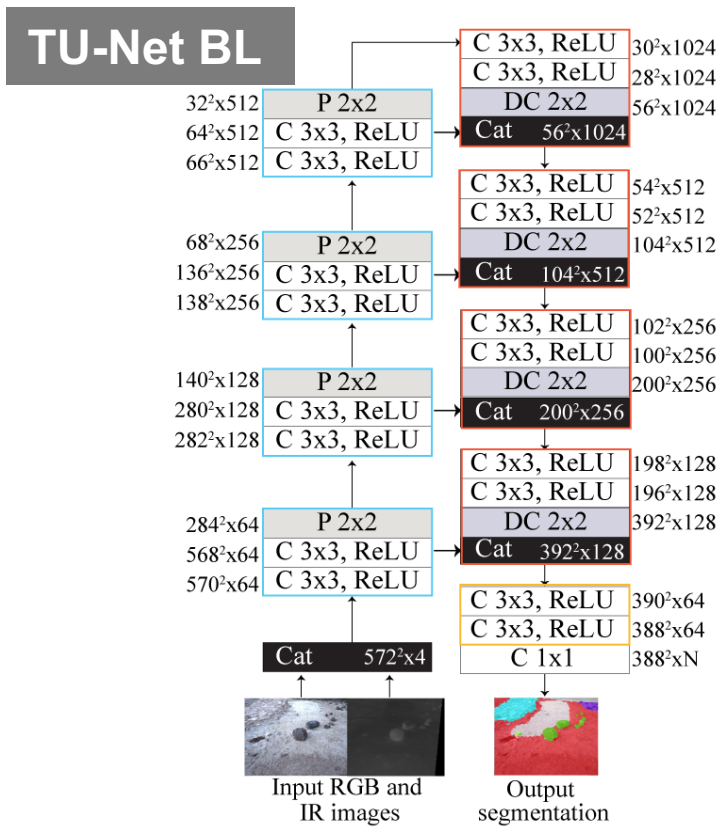
Cat: concatenate
C: convolution
ReLU: rectified linear unit
P: pooling
DC: deconvolution

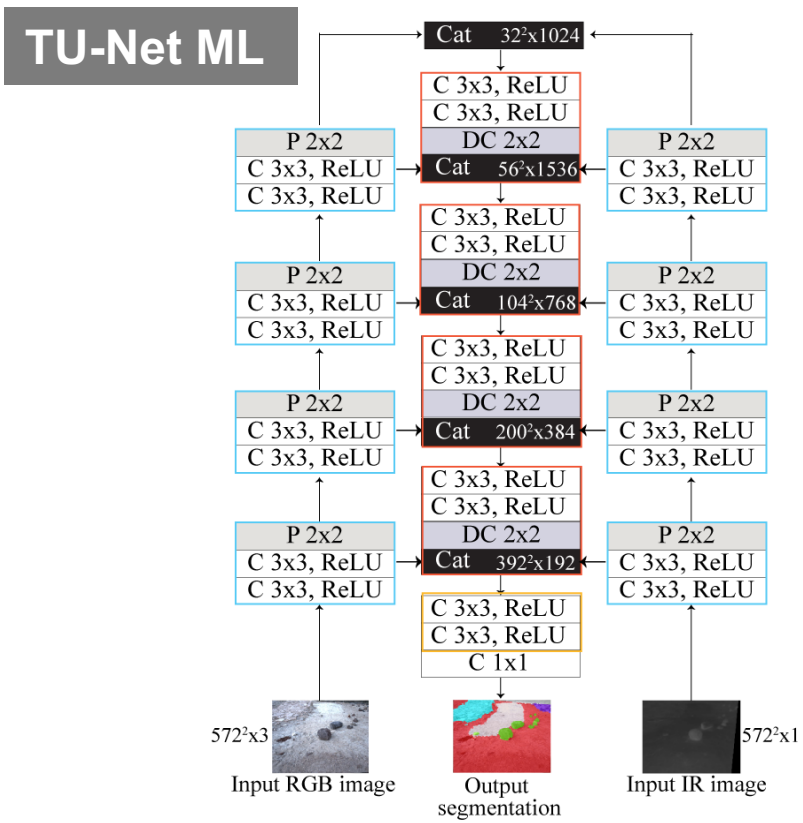


Loss function for terrain classification $\mathcal{L}_{CE}$:
Cross entropy **from annotation images**

TU-Net (Two U-Net)

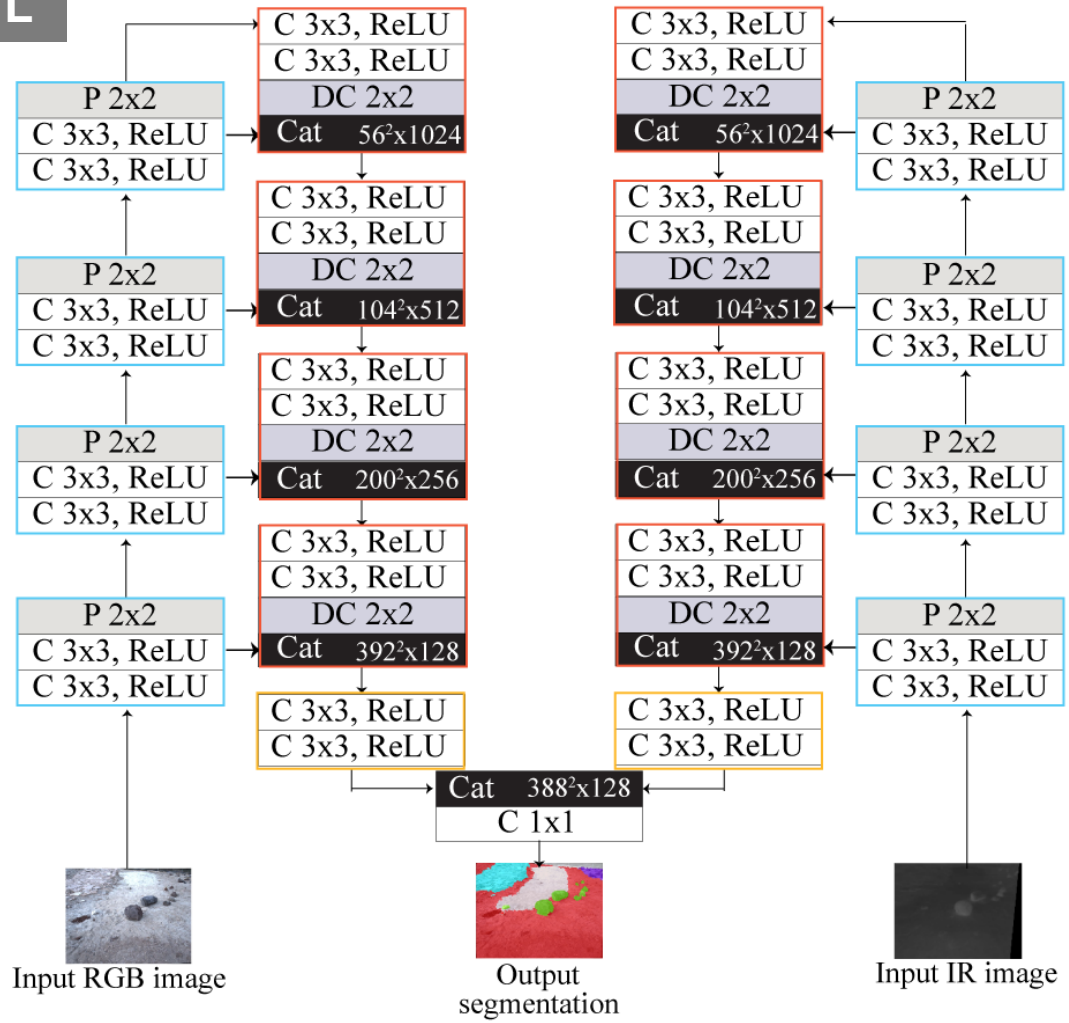
- Three different architectures: fusion of visible and thermal images at different level
  - **TU-Net BL (bottom layer): fusion of local information**
  - **TU-Net ML (middle layer): fusion of global information**
  - **TU-Net TL (top layer): fusion of feature-level information**





TU-Net (Two U-Net)

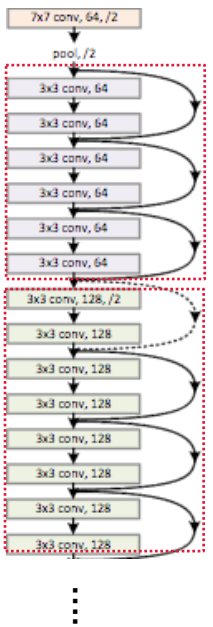
TU-Net TL



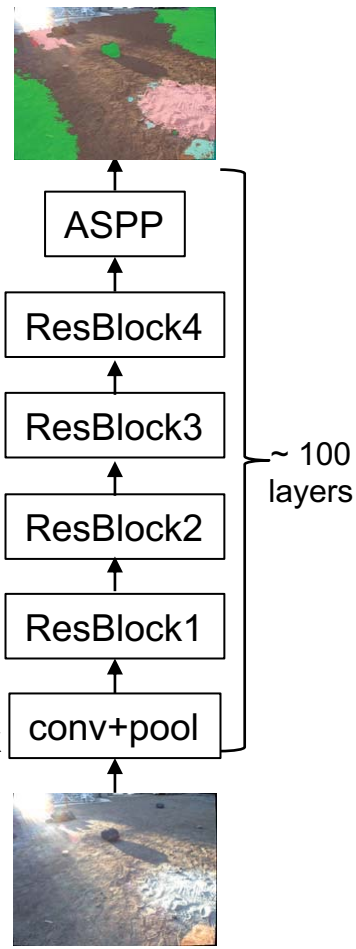
TDeepLab (Two DeepLab)

[1] L. Chen, et al., “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”, arXiv 2016

- TDeepLab is based on DeepLab v2 (Backbone is ResNet-101 v1)
  - 2 architectures: **TDeepLab BL and TDeepLab TL**

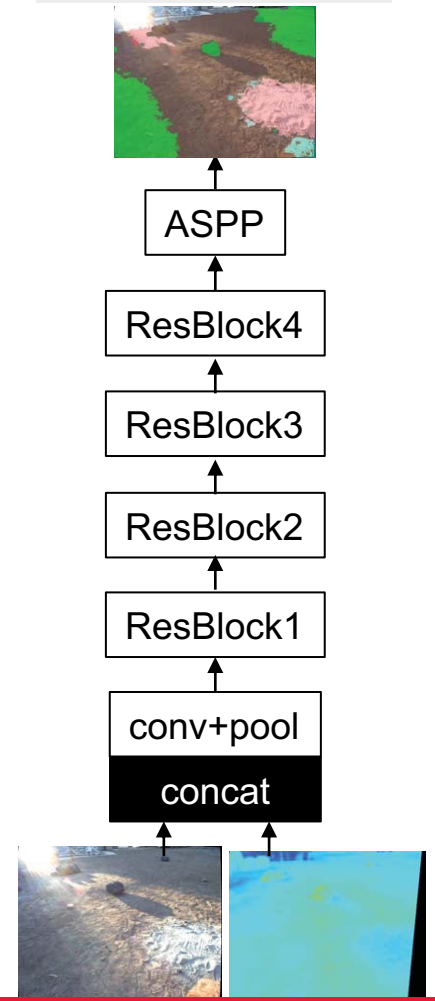


DeepLab (network for single modal)

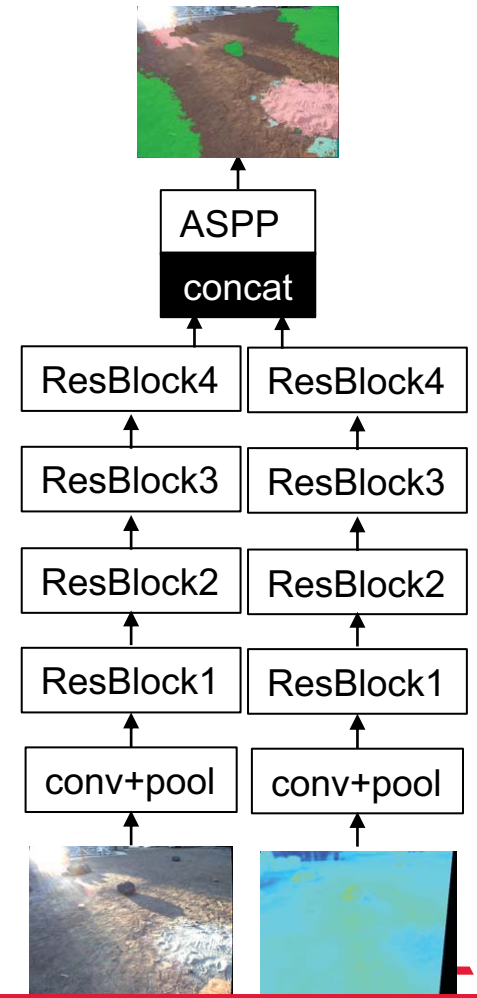


ResBlock: Residual Block
ASPP: Atrous Spatial Pyramid Pooling

TDeepLab BL



TDeepLab TL

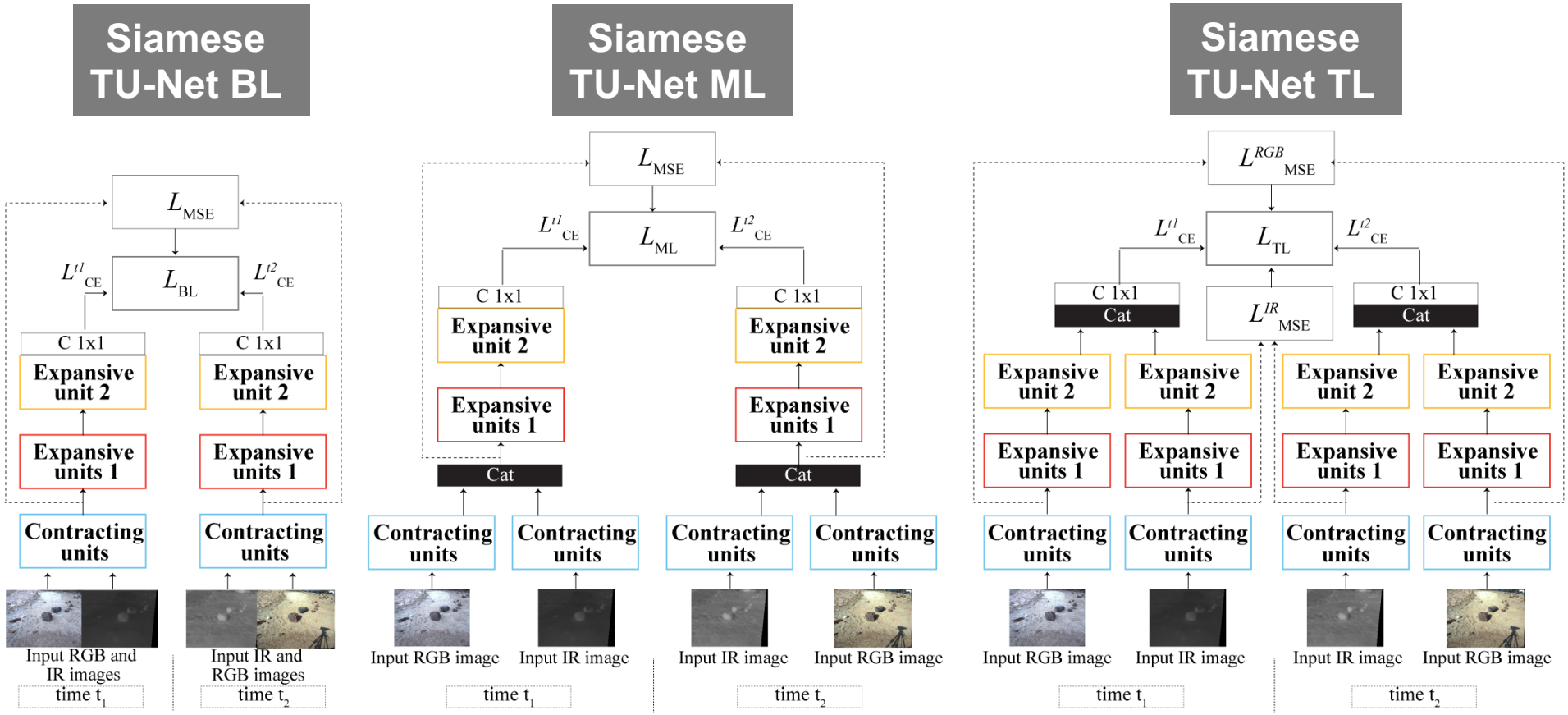


Siamese-based TU-Net and TDeepLab

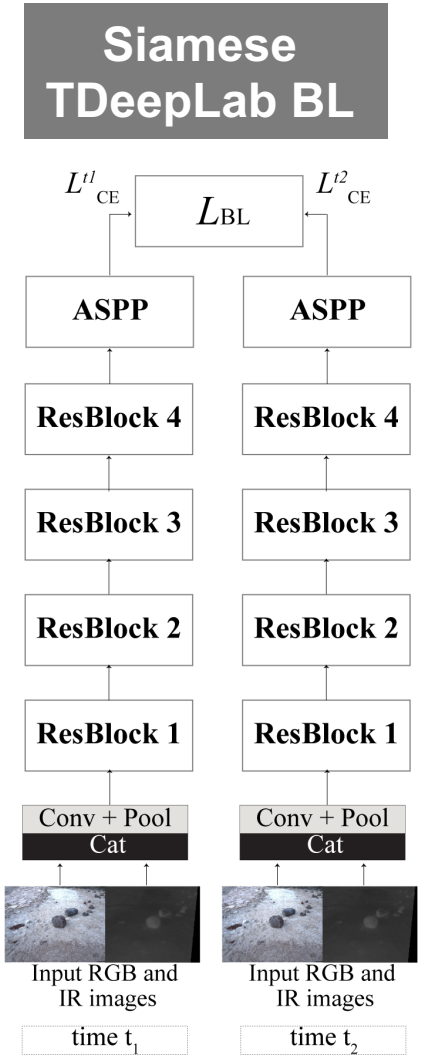
- TU-Net and TDeepLab can implicitly learn ***implicitly*** illumination changes with images captured at various time of a day

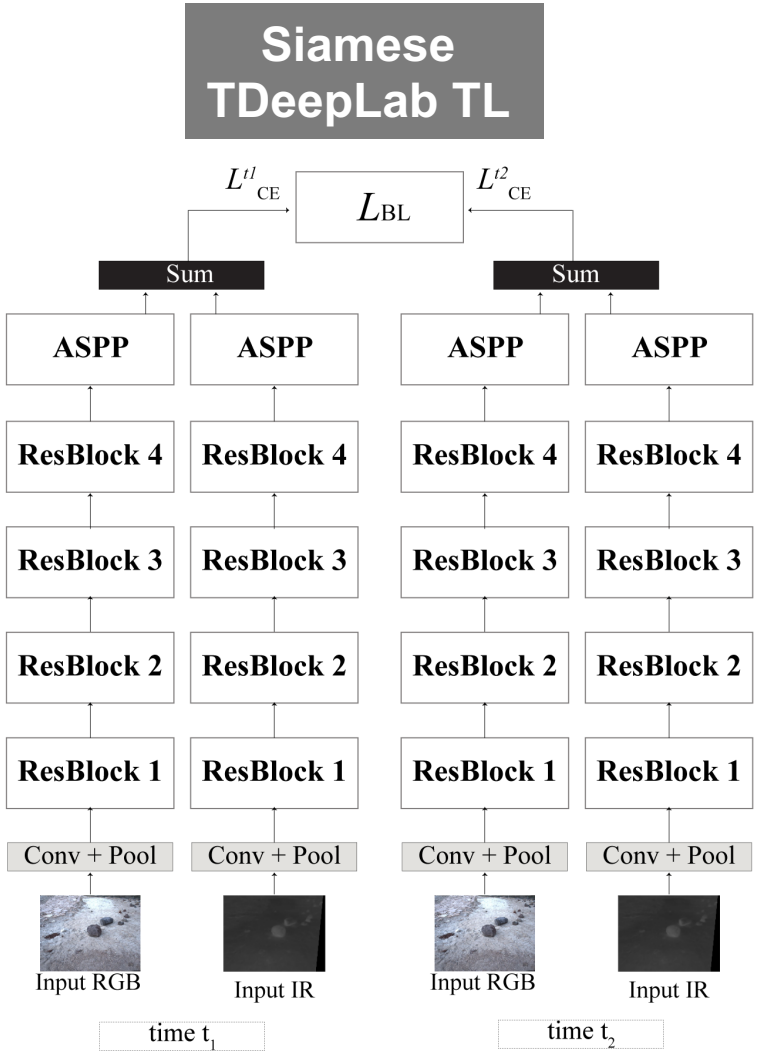
- To make it ***explicitly***, we integrate a Siamese approach into TU-Net and TDeepLab
  - 3 architectures of Siamese-based TU-Net
    - Siamese TU-Net BL, Siamese TU-Net ML, Siamese TU-Net TL
  - 2 architectures of Siamese-based TDeepLab
    - Siamese TDeepLab BL, Siamese TDeepLab TL

Siamese-based TU-Net



Siamese-based TDeepLab



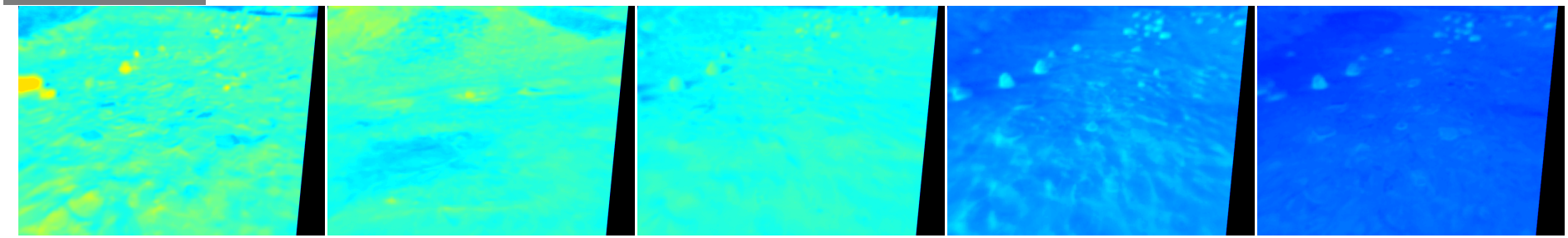


Experiment setting

- Dataset collection
  - Every 1 hour from 10 am till 5 pm on Nov 17th 2017, 52 images at each time
  - RGB camera (FLIR Grasshopper 5M) and thermal camera (FLIR AX65)
    - RGB images include challenging illumination changes

RGB images

IR images

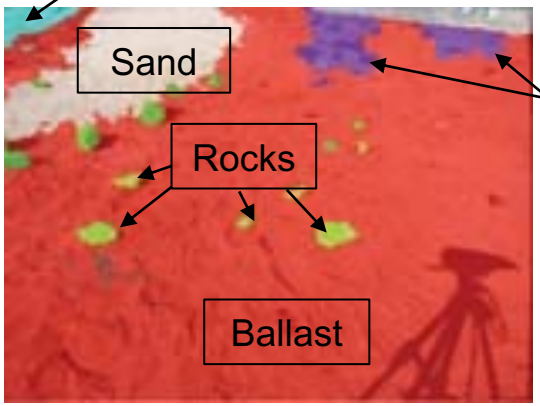


10 am | 12 pm | 2 pm | 4 pm | 5 pm

Experiment setting

- Manually annotated images into 6 categories
  - Sand, soil, rocks, bedrock, rocky terrain, and ballast

Captured RGB image



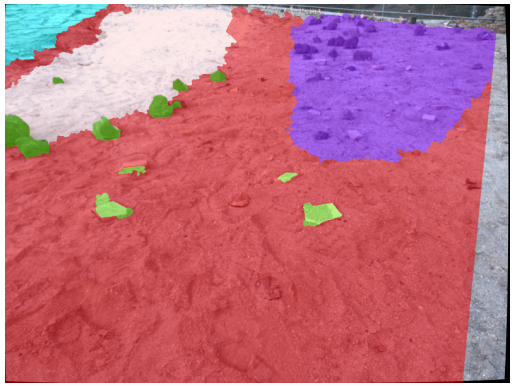
Ground truth annotated image

  - At each time 52 images are separated into 3 dataset
    - 50 % for training, 25 % for evaluation, and 25% for test dataset

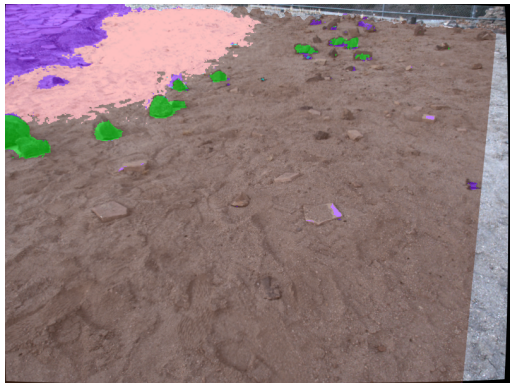
- Training a model
  - TU-Net is trained from a scratch
  - TDeepLab is finetuned (initial parameters are trained with COCO)

Exp 1. Train, evaluate, and test (all 5pm images)

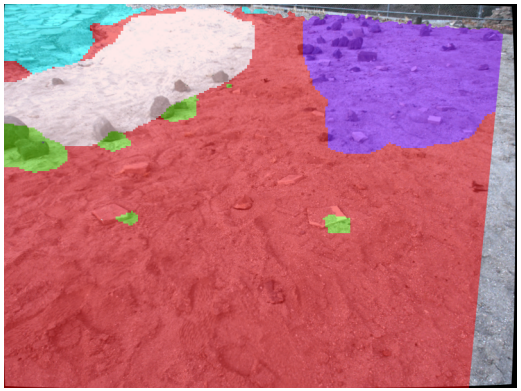
- Visualization of example results



Ground truth annotated image



Estimated result by TU-Net TL



Estimated result by TDeepLab TL

	Pixel accuracy	Mean accuracy	Mean IoU*	Frequency weighted IoU*
U-Net (RGB)	0.632	0.481	0.239	0.512
TU-Net BL (RGB + IR)	0.678	**0.529**	0.283	0.552
TU-Net ML (RGB + IR)	0.684	0.500	0.262	0.565
TU-Net TL (RGB + IR)	**0.727**	0.497	**0.304**	**0.592**
DeepLab (RGB)	0.871	0.708	0.599	0.781
TDeepLab BL (RGB+IR)	0.859	0.772	0.610	0.792
TDeepLab TL (RGB+IR)	**0.911**	**0.820**	**0.679**	**0.848**

*IoU (Intersection over Union)

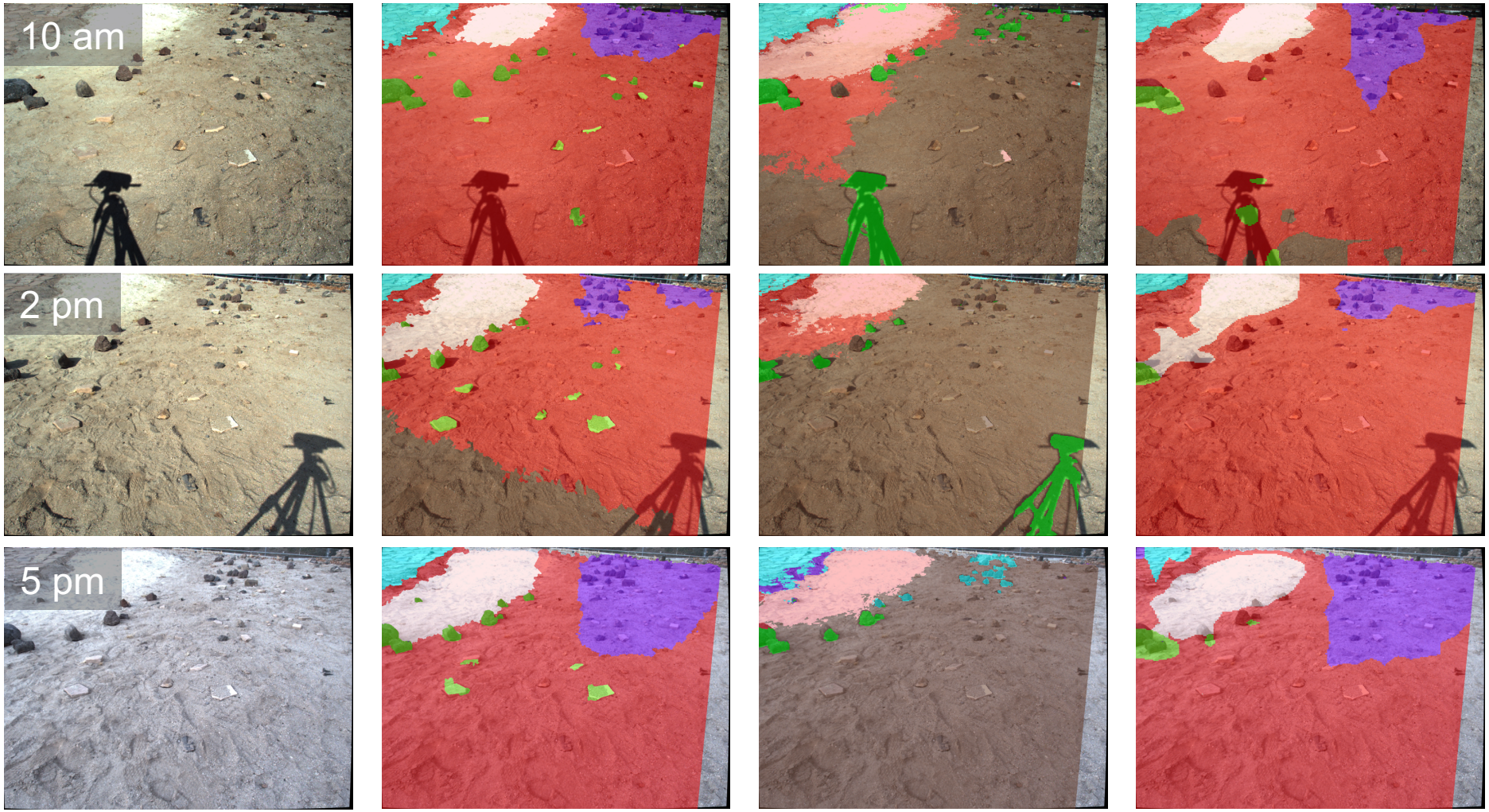
Exp 2. Train, evaluate, and test (all 10am-5pm images)

- Siamese-based TU-Net works better than TU-Net
  - Siamese TU-Net TL calculates a loss function of IR independently from the one of RGB
    - IR information without RGB shows some ambiguity
  - Both Siamese TU-Net BL and ML combines a loss function of IR and RGB
- On the other hand, Siamese-based TDeepLab is not effective
  - One of the reason might be over-trained

	Pixel accuracy	Mean accuracy	Mean IoU*	Frequency weighted IoU*
TU-Net TL	0.581	0.435	0.211	0.447
Siamese TU-Net BL	0.612	**0.542**	**0.307**	0.491
Siamese TU-Net ML	**0.620**	0.534	0.303	**0.505**
Siamese TU-Net TL	0.563	0.517	0.257	0.463
TDeepLab TL	**0.911**	**0.820**	**0.679**	**0.848**
Siamese TDeepLab BL	0.803	0.665	0.482	0.696
Siamese TDeepLab TL	0.864	0.731	0.600	0.771

Exp 3. Train and evaluate (2pm-5pm), and test (10am-1pm)

- Visualization of example results



Original image | Annotated image | Siamese TU-Net ML | TDeepLab TL

Exp 3. Train and evaluate (2pm-5pm), and test (10am-1pm)

- Results of Siamese TU-Net ML

Test data	Pixel accuracy	Mean accuracy	Mean IoU*	Frequency weighted IoU*
10am – 1pm	0.633	0.461	0.279	0.531
2pm – 5pm	0.687	0.529	0.374	0.559

- Results of TDeepLab TL

Test data	Pixel accuracy	Mean accuracy	Mean IoU*	Frequency weighted IoU*
10am – 1pm	0.861	0.816	0.625	0.772
2pm – 5pm	0.886	0.786	0.653	0.806

Conclusions

- We proposed novel deep learning-based terrain classification methods robust to illumination variations, called TU-Net and TDeepLab and those Siamese-based methods

- Future work
  - Compare the proposed approaches with AdapNet++
  - Integrate depth information

Backup

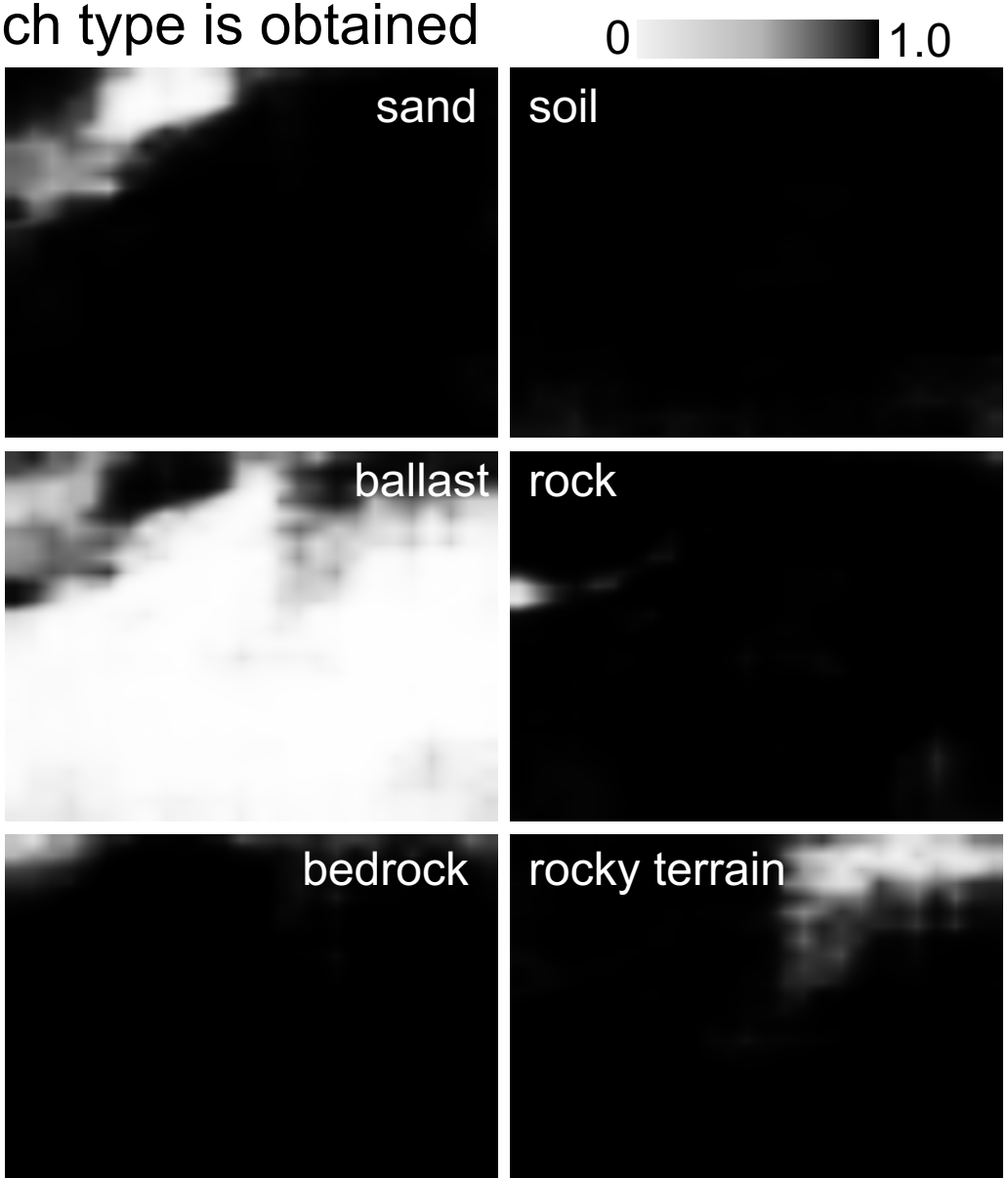
TDeepLab

- Probability of each type is obtained

Result by TDeepLab

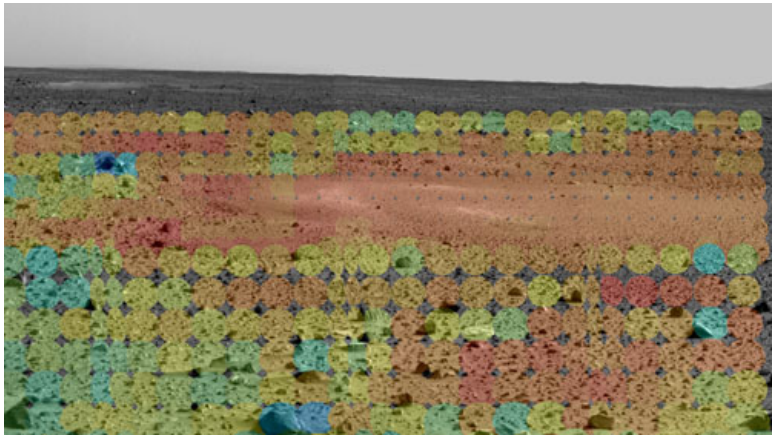
Probabilities
sand: 0.35
ballast: 0.65

It can estimate how many types / the proportion of terrain types



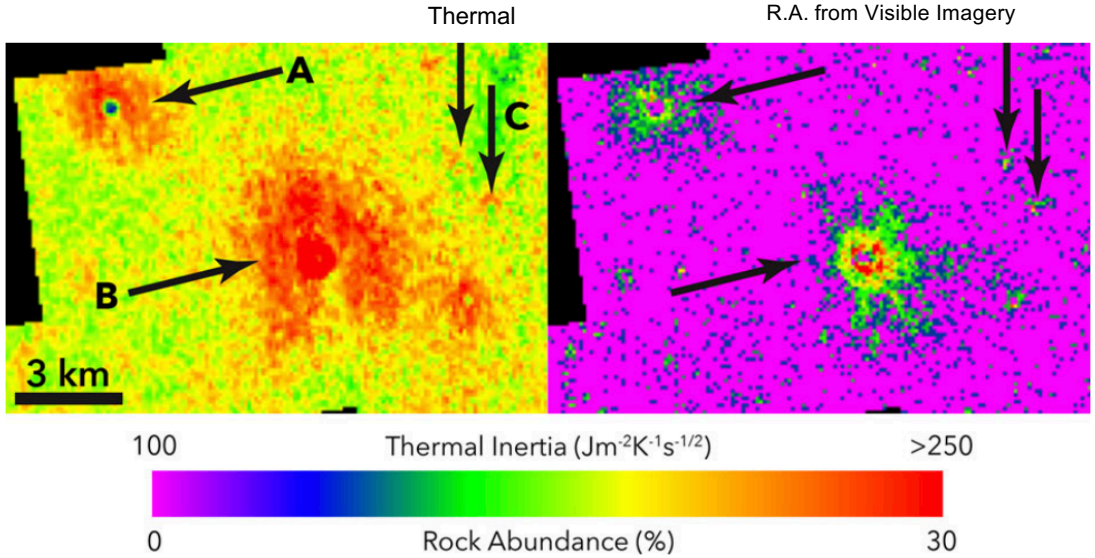
Benefits of IR images

- Remote measurements of mineralogy and thermophysical properties of the scene



Colorized mini-TES image (Spirit). The different colored circles show a spectrum of soil and rock temperatures

- Estimate rock size from orbiters
  - Useful for future landing site selection



Golombek et al. 2018