INSTITUTE OF COMPUTER SCIENCE

4D Convolutional Neural Networks for Multi-Spectral and Multi-Temporal Remote Sensing Data Classification

M. Giannopoulos, G. Tsagkatakis, P. Tsakalides

Signal Processing Lab, Institute of Computer Science-Foundation for Research and Technology (FORTH)-Hellas, Crete, Greece

Computer Science Department, University of Crete, Crete, Greece



Agenda

PART A: Motivation

a. Remote Sensing Data Classification

b. Multi-Spectral & Multi-Temporal Imaging

PART B: Proposed Method

a. 4D-Convolutional Neural Networks

b. Multi-Temporal Remote Sensing Land-Cover Classification

c. 4D Model Architecture

PART C: Experimental Evaluation

a. Dataset Description-Dataset Visualization

b. Experimental Setup

c. Experimental Results

PART D: Conclusions and Future Directions

Acknowledgements

Agenda

PART A: Motivation

a. Remote Sensing Data Classification

b. Multi-Spectral & Multi-Temporal Imaging

PART B: Proposed Method

a. 4D-Convolutional Neural Networks

b. Multi-Temporal Remote Sensing Land-Cover Classification

c. 4D Model Architecture

PART C: Experimental Evaluation

a. Dataset Description-Dataset Visualization

b. Experimental Setup

c. Experimental Results

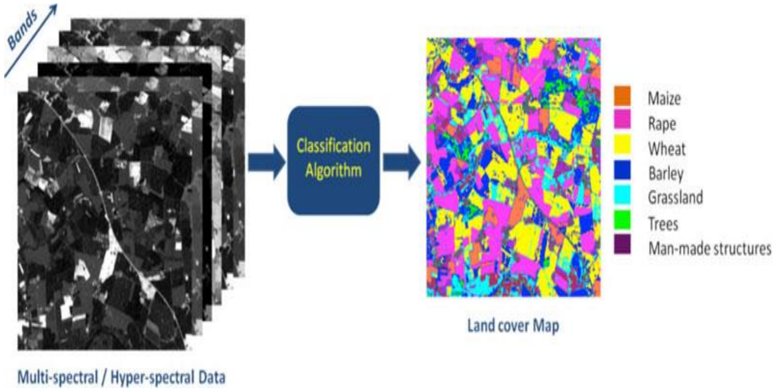
PART D: Conclusions and Future Directions

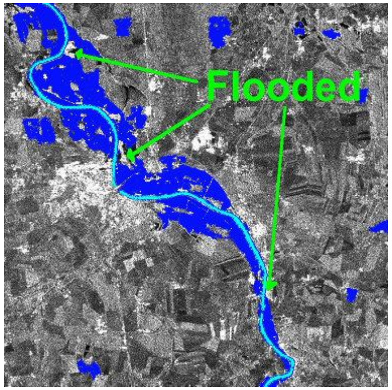
Acknowledgments

Remote Sensing Data Classification

High-resolution satellite sensors → **Periodic data collection**

- Massive <u>high-dimensional observations</u> acquired across multiple time instances
- Land-cover classification, semantic segmentation
- Flood detection, crop classification



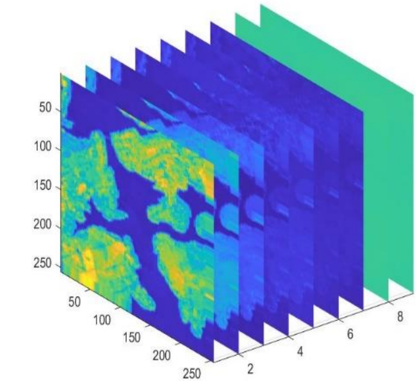


Efficient way of capturing multi-dimensional dependencies & correlations!
Avoid any information loss!

MS & MT Imaging

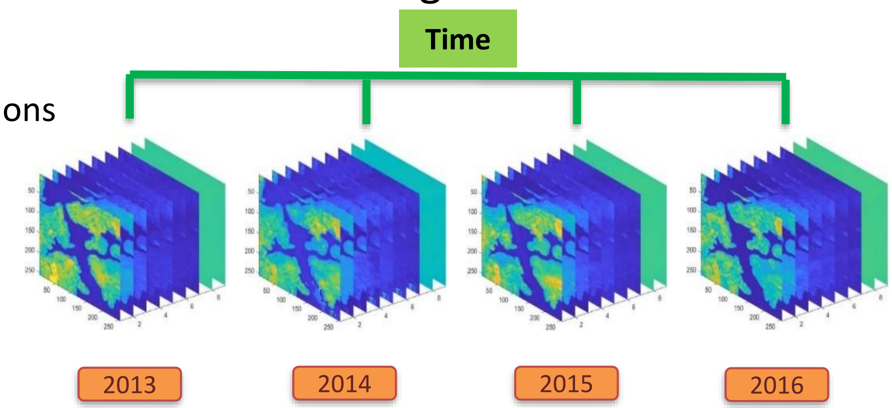
- **<u>Multi-Spectral (MS) Imaging</u>**
  - ➢ Acquire info across the EM spectrum
  - ➢ Spectrum of light that is scattered by some materials on Earth's surface
  - ➢ Usually ~3-12 spectral bands
  - ➢ 2 spatial & 1 spectral dimensions

3D MS data-cube



- **<u>Multi-Temporal (MT) Imaging</u>**
  - ➢ Acquire remotely sensed data from >1 time period
  - ➢ Information about how our world is changing
  - ➢ Provide the tools to monitor land use and land cover change
  - ➢ Usually time-series of MS images
    - ❑ 2 spatial, 1 spectral & 1 temporal dimensions

4D MS-MT data-cube



Agenda

PART A: Motivation

a. Remote Sensing Data Classification

b. Multi-Spectral & Multi-Temporal Imaging

PART B: Proposed Method

a. 4D-Convolutional Neural Networks

b. Multi-Temporal Remote Sensing Land-Cover Classification

c. 4D Model Architecture

PART C: Experimental Evaluation

a. Dataset Description-Dataset Visualization

b. Experimental Setup

c. Experimental Results

PART D: Conclusions and Future Directions

Acknowledgements

4D-CNNs (I)

- **<u>Goal</u>**: Exploit spatial, spectral and temporal information at the same time
- <u>How</u>: Stacking multiple sequences of 3D convolutions along the last dimension

$$y_{k,l,m,n} = f\left(\sum_{c}^{C_{in}} \sum_{s=0}^{S-1} \sum_{t=0}^{T-1} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} w_{i,j,t,s} x_{c,(k+i)(l+j)(m+t)(n+s)} + b_{i,j,t,s}\right)$$

$$= f\left(\sum_{s=0}^{S-1} \left[\sum_{c}^{C_{in}} \sum_{t=0}^{T-1} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} w_{i,j,t,s} x_{c,(k+i)(l+j)(m+t)(n+s)}\right] + b_{i,j,t,s}\right)$$

$$= f\left(\sum_{s=0}^{S-1} C_{3D} + b_{i,j,t,s}\right)$$

Involved Parameters

- $y_{k,l,m,n}$: Convolved output neuron at position (k, l, m, n)
- $f(.)$: Activation function
- $w_{i,j,t,s}$: Value of the kernel connected to the current feature map at position (i, j, t, s)
- $x_{c,(k+i)(l+j)(m+t)(n+s)}$: Value of the input neuron at channel c
- $b_{i,j,t,s}$: Bias of the computed feature map
- H: Height, W: Width, T: Temporal length, S: Spectral bands
- C_{in}: Number of original channels/feature maps of previous layer
- C_{3D}: 3D-convolution operator

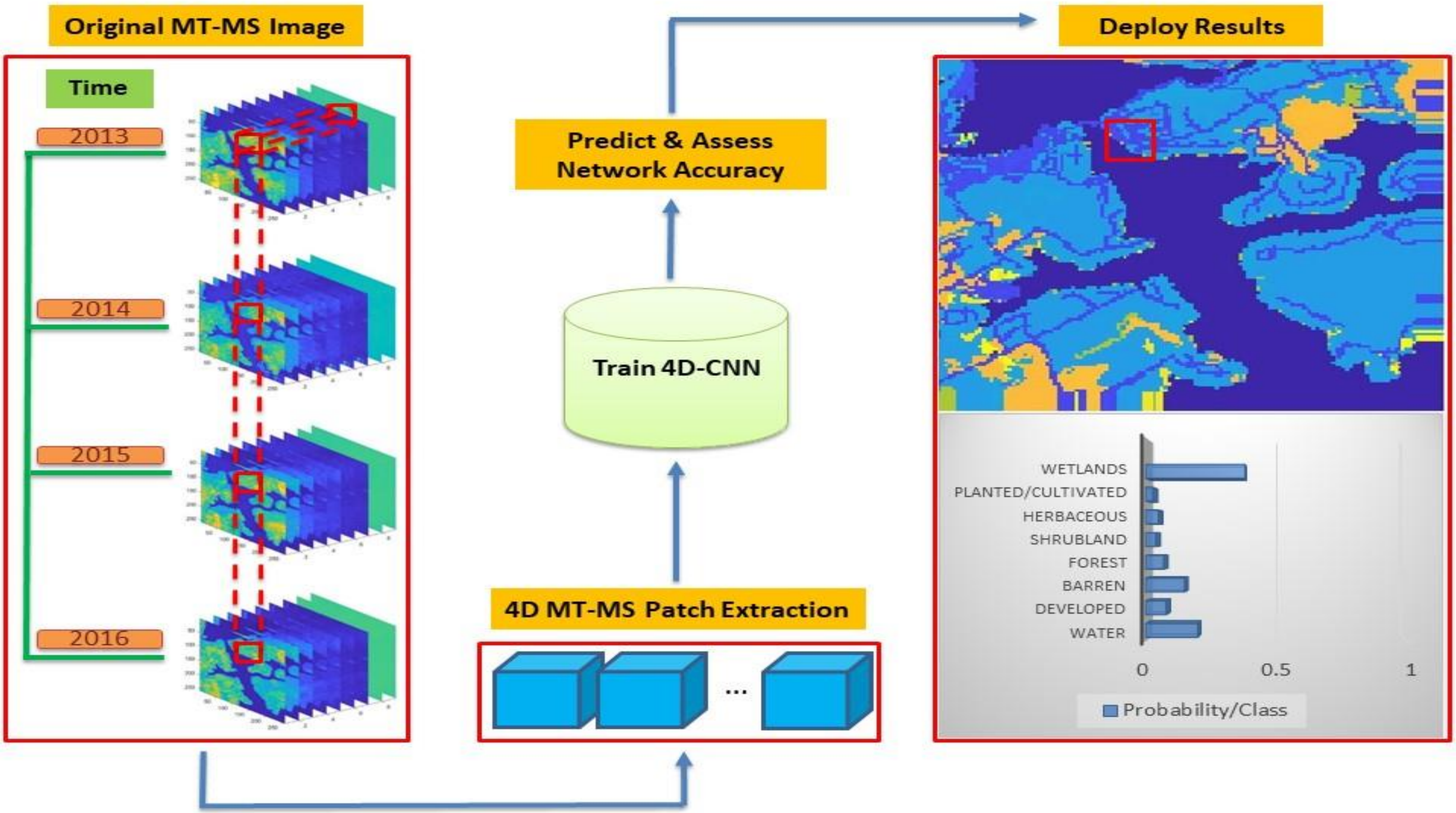
4D-CNNs (II)

Why going stacked

- **Fast** GPU-based **implementation** using Tensorflow primitives
- <u>Is it feasible</u>? Yes
  - Convolution: Linear operation → Summation order can change
  - Implementation: <u>Further re-arrangement of sums' indices [1-2]</u> → Not separable convolution

[1] A. Myronenko et. al, "4D CNN for Semantic Segmentation of Cardiac Volumetric Sequences", in Proceedings of the *International Workshop on Statistical Atlases and Computational Models of the Heart*, Lima, Peru, 4 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 72-80.
[2] M. Giannopoulos, G. Tsagkatakis and P. Tsakallides "4D U-Nets for Multi-Temporal Remote Sensing Data Classification", *Remote Sensing*, Vol. 14, No. 3, Multidisciplinary Digital Publishing Institute, 28 January 2022 .

MT-RS Land-Cover Classification



4D Model Architecture

Fully Convolutional (FC) network

- No pooling layers:
  - ➢ No need for considering scaling/translation factors
- FC nets perform remarkably in similar tasks (i.e. Hyper-Spectral pixel-level classification [3,4])

Model Topology

- Stacks of (Convolutional-Batch Normalization-ReLU) layers
- Kernel-size: Equal to 3 across every dimension
  - ➢ Padding="same"
- Loss function: Categorical cross-entropy
- Optimizer: Adam
  - ➢ Learning-rate=0.0001
  - ➢ $\beta_1 = 0.9, \beta_2 = 0.999$
- Batch-size: 128
- Epochs: 100

[3] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 4959-4962.
[4] M. Giannopoulos, G. Tsagkatakis and P. Tsakallides, "On the Impact of Tensor Completion in the Classification of Undersampled Hyperspectral Imagery", in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1975-1979 .

Agenda

PART A: Motivation

a. Remote Sensing Data Classification
b. Multi-Spectral & Multi-Temporal Imaging

PART B: Proposed Method

a. 4D-Convolutional Neural Networks
b. Multi-Temporal Remote Sensing Land-Cover Classification
c. 4D Model Architecture

PART C: Experimental Evaluation

a. Dataset Description-Dataset Visualization
b. Experimental Setup
c. Experimental Results

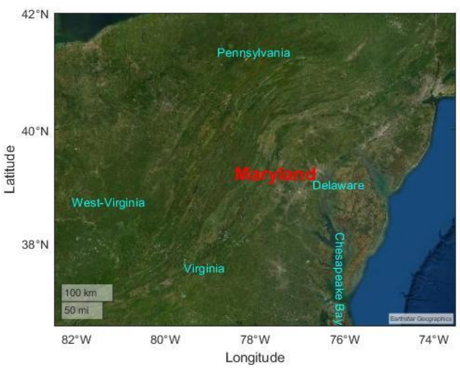
PART D: Conclusions and Future Directions

Acknowledgments

Dataset Description

IEEE GRSS Data Fusion Contest dataset
2250 image tiles over Maryland-State, USA

- Landsat-8 MS imagery, collected once a year through 2013-2016
  - 4000x4000x9x4 measurements (spatial, spatial, spectral, temporal)
- USGS NLCD labels of 2016
  - 15 different land-cover classes



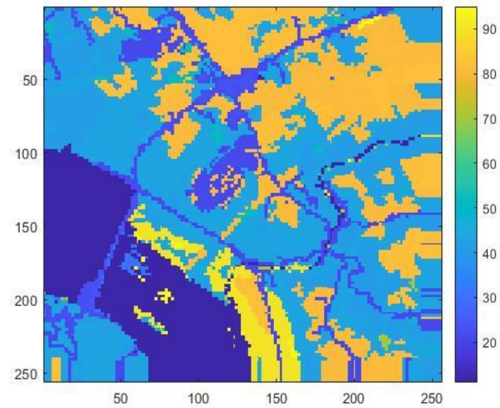
- **Pre-processing steps**:
  - Selection of a tile with no undefined label values
  - Spatial padding, until spatial size reaches (4096,4096) pixels
  - Spatial subsampling every 16 pixels, until spatial size reaches (256,256) pixels
  - Spatial patch-extraction around each pixel, across all spectral bands & time instances
  - Created data: 65536 samples, of size (p,p,9,4) each
- Dataset split: Uniformly at random
  - Training-set: 60%, Validation-set: 20%, Test-set: 20%

Dataset Visualization

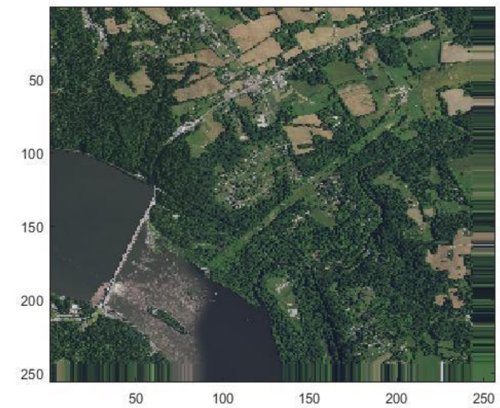
Labels' names

Class ID	Class Value	Class Name
1	11	Open Water
2	21	Developed, Open Space
3	22	Developed, Low Intensity
4	23	Developed, Medium Intensity
5	24	Developed High Intensity
6	31	Barren Land (Rock/Sand/Clay)
7	41	Deciduous Forest
8	42	Evergreen Forest
9	43	Mixed Forest
10	52	Shrub/Scrub
11	71	Grassland/Herbaceous
12	81	Pasture/Hay
13	82	Cultivated Crops
14	90	Woody Wetlands
15	95	Emergent Herbaceous Wetlands

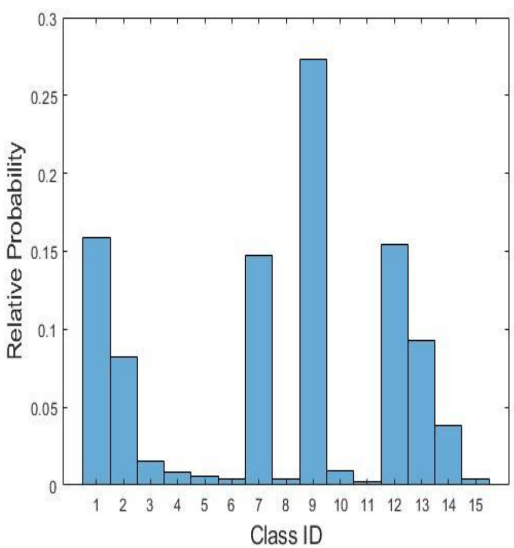
Labels' overview



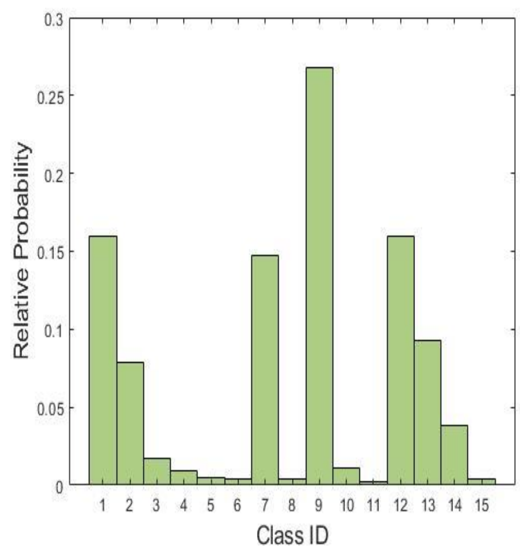
Labels' overview-RGB



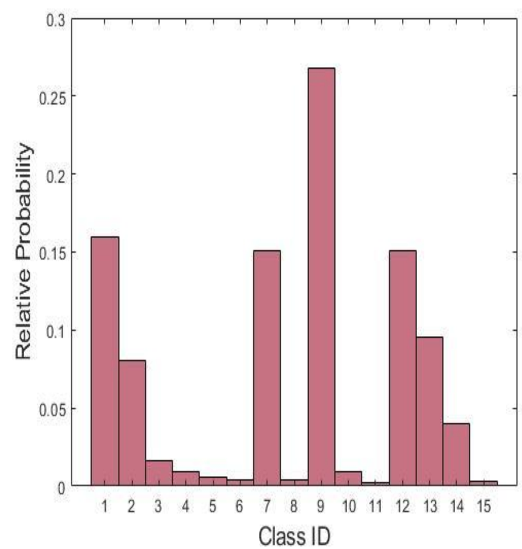
Labels' distribution: Training Set



Labels' distribution: Validation Set



Labels' distribution: Test Set



Experimental Setup

Cross-Validated Parameters

- #Stack-of-layers in each architecture (i.e. 2, 3, 4)
- Spatial patch-size of training samples (i.e. 5x5, 7x7)

Employed Platforms

- Software: Python, Tensorflow, Keras
- Hardware: NVIDIA Quadro P4000 (8Gb RAM)

Deep Learning Models

- 2D-CNN: Exploit spatial info
- 3D-CNN-T: Exploit spatial & temporal info
- 3D-CNN-S: Exploit spatial & spectral info
- **4D-CNN (Proposed)**: Exploit spatial & spectral & temporal info

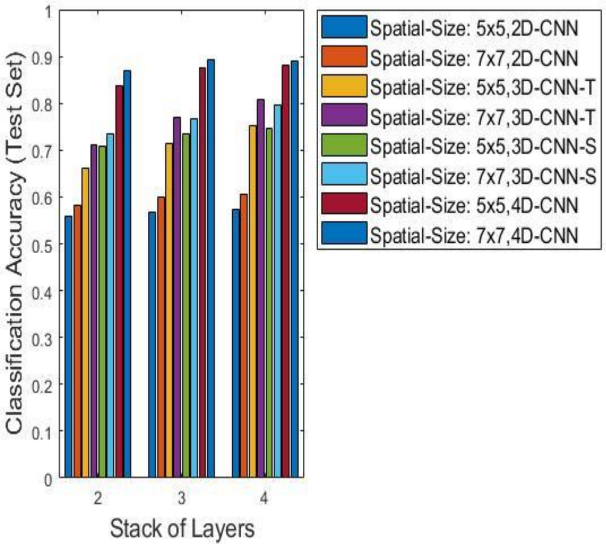
Comparison ML and S-o-t-A Models

- Support Vector Machines (Gaussian Kernel)
- k-Nearest Neighbors (k=5)
- 3D-CNN [8]

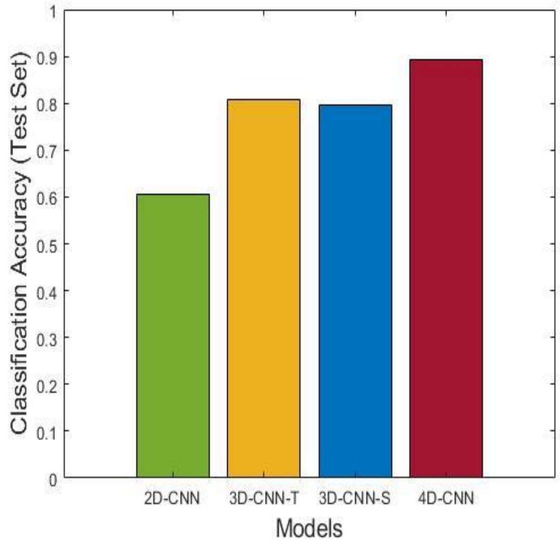
[8] Shunping Ji, Chi Zhang, Anjian Xu, Yun Shi, and Yulin Duan, "3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images," *Remote Sensing*, Vol. 10, No. 1, pp. 75, 2018.

4D-CNN Architectures Parameter Tuning

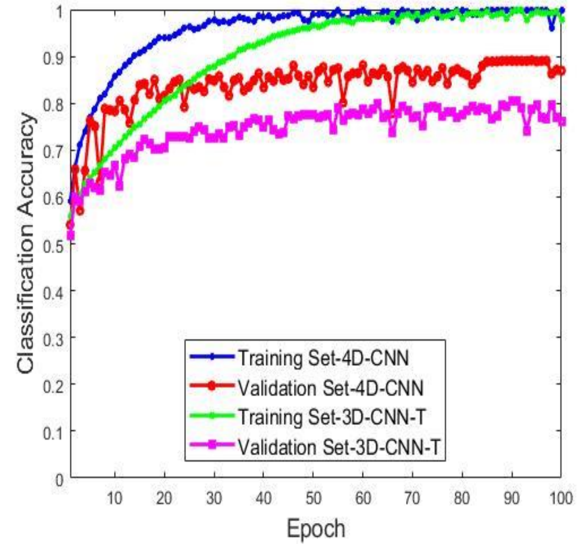
All trained CNN models



Best trained CNN models



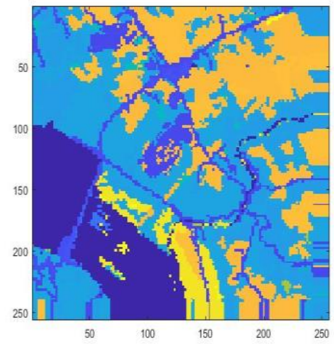
Top-2 trained CNN models



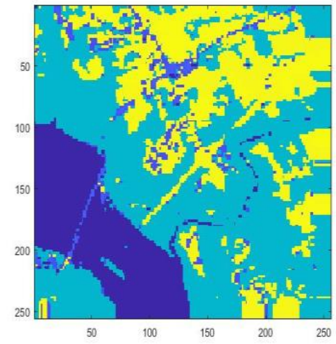
- **Goal**: Assess the performance of the CNN models relative to:
  - The number of stack-of-layers they consist of
  - The spatial patch-size they are trained with
- 4D-CNN model achieves an accuracy improvement of up to 8.36% over the second best model (89.16% vs 80.80%)!
- 4D-CNN model converges faster to higher classification accuracy value
- 4D-CNN model less prone to over-fitting issues

Comparison to S-o-t-A & ML Methods

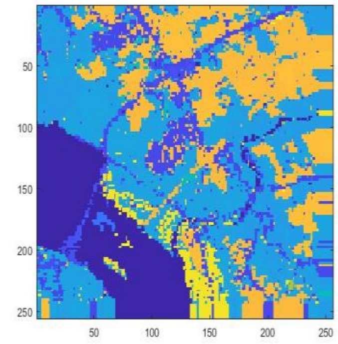
Actual labels



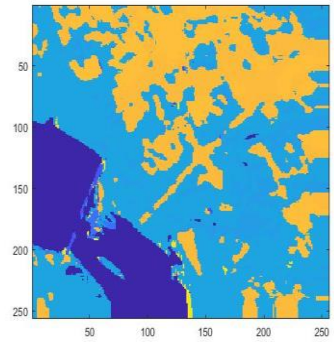
Predicted labels-SVM-Gaussian



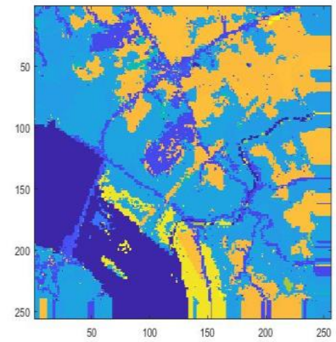
Predicted labels-k-NN



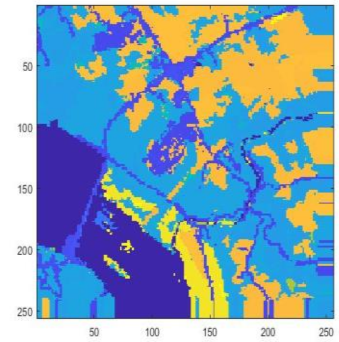
Predicted labels-3D-CNN [8]



Predicted labels-3D-CNN-T



Predicted labels-4D-CNN



Model	Accuracy	Time	F1-Score
k-NN	0.7443	0.000	0.5747
SVM-Gaussian	0.5945	2.006	0.4995
3D-CNN [8]	0.7097	6.003	0.5090
2D-CNN	0.6049	4.928	0.3624
3D-CNN-T	0.8080	14.558	0.6509
3D-CNN-S	0.7953	27.435	0.6552
4D-CNN	**0.8916**	**107.619**	**0.7796**

- **Goal**: Compare best CNN models with S-o-t-A and ML methods
- 4D-CNN model achieves an F1-Score improvement of up to 12.44% over the second best model (77.96% vs 65.52%) → Class-imbalance robustness
- 4D-CNN model is ~7 slower than the second best model (~1.8hrs vs ~15mins)
- 4D-CNN model ends up with clear segmented regions w.r.t. ground-truth labels

Agenda

PART A: Motivation

a. Remote Sensing Data Classification

b. Multi-Spectral & Multi-Temporal Imaging

PART B: Proposed Method

a. 4D-Convolutional Neural Networks

b. Multi-Temporal Remote Sensing Land-Cover Classification

c. 4D Model Architecture

PART C: Experimental Evaluation

a. Dataset Description-Dataset Visualization

b. Experimental Setup

c. Experimental Results

PART D: Conclusions and Future Directions

Acknowledgments

Conclusions & Future Directions

- **Efficient classification** of MT-RS data via 4D-CNNs
- **Clear improvements** over lower-dimensional CNNs, machine learning and state-of-the-art methods
- **Robustness** towards class-imbalance regimes

- **4D-Convolutional Neural Networks**
  - Effective exploitation of higher-order correlations without any information loss
  - End-to-end learning of spatio-spectro-temporal features at the same time
  - Computational burden not prohibitive

- **Future Work**
  - Ameliorate implementation→ Faster models

Acknowledgements

spl.edu.gr

SPL-ICS-FORTH-Hellas, Crete, Greece
CSD, UoC, Crete, Greece

mgiannop@ics.forth.gr
greg@ics.forth.gr
tsakalid@ics.forth.gr

This work was funded by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Innovation (GSRI), under the **HFRI Faculty Grant 1725 (V4-ICARUS)**, and by the **Calchas project (contract no.842560)** of the H2020 Framework Program of the European Commission