2° Workshop on Reproducible Research in Pattern Recognition (RRPR2018) 20 August 2018

On the Reproducibility of Deep Convolutional Neural Networks Approaches

Gabriele Piantadosi, Stefano Marrone, and Carlo Sansone
{gabriele.piantadosi, stefano.marrone, carlo.sansone}@unina.it
DIETI - University of Naples Federico II, Italy

Reproducibility issues in Deep Learning

NVIDIA CUDA Deep Neural Network (cuDNN) Developer Guide

2.5. Reproducibility (determinism)

By design, most of cuDNN's routines from a given version generate the same bit-wise results across runs when executed on GPUs with the same architecture and the same number of SMs. However, bit-wise reproducibility is not guaranteed across versions, as the implementation of a given routine may change. With the current release, the following routines do not guarantee reproducibility because they use atomic operations:

- `cudnnConvolutionBackwardFilter` when `CUDNN_CONVOLUTION_BWD_FILTER_ALGO_0` or `CUDNN_CONVOLUTION_BWD_FILTER_ALGO_3` is used
- `cudnnConvolutionBackwardData` when `CUDNN_CONVOLUTION_BWD_DATA_ALGO_0` is used
- `cudnnPoolingBackward` when `CUDNN_POOLING_MAX` is used
- `cudnnSpatialTfSamplerBackward`

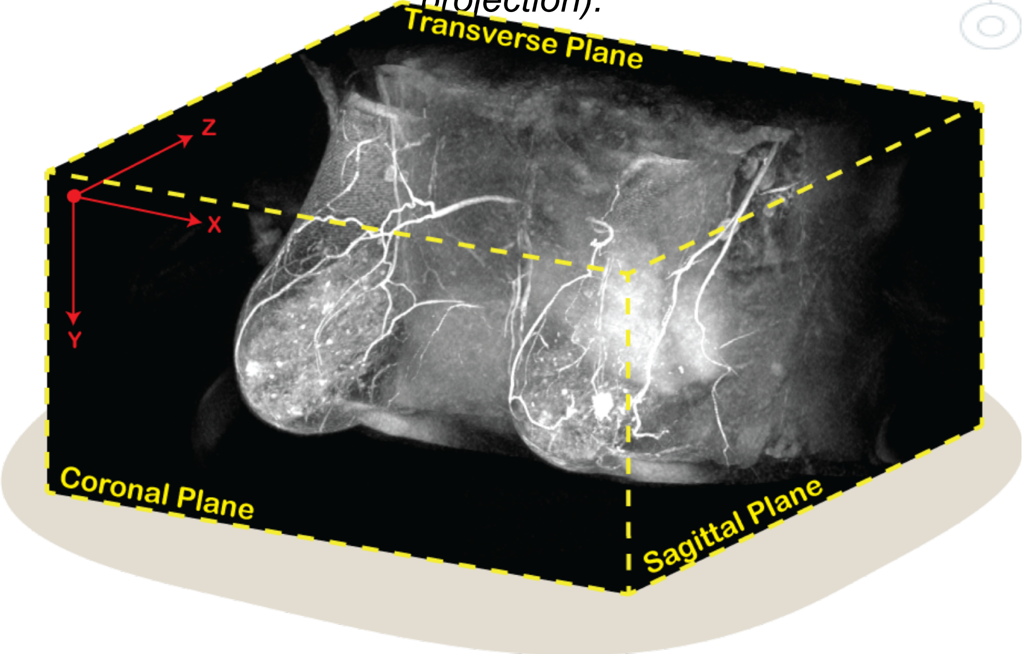
Intuition: the reproducibility issue can be shifted from a strictly combinatorial problem to a statistical one, in order to validate the model robustness and stability more than its perfect outcomes predictability

Case Study: U-NET for Breast Tissues Segmentation

Dynamic Contrast Enhanced-Magnetic Resonance Imaging **(DCE-MRI)**

- ✓ Important complementary diagnostic methodology
- ✓ Minimal invasiveness (non-ionizing radiation)
- ✓ Produces 4D volumes (3 spatial + 1 temporal)
- ✓ The paramagnetic contrast agent enhance vascularization
- ✓ Suited for under-forty patients
- ✗ Huge amount of data
- ✗ Long acquisition time.
- ✗ Involuntary patient movements

A 3D MRI volume [UP] (pre-contrast volume in maximum intensity projection).



Case Study: U-NET for Breast Tissues Segmentation

Computer Aided Detection and Diagnosis systems **(CAD)**[1]

Volume Extraction → **BreastMask Extraction** → **Preprocessing** → **ROI Detection** → **ROI Classification**

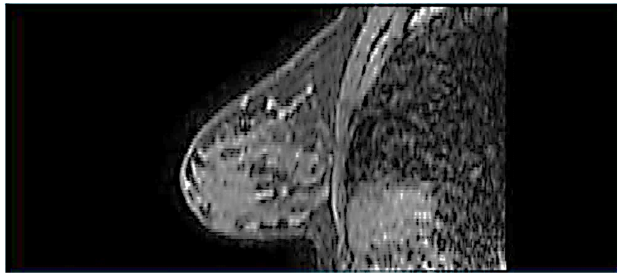
- **Volume Extraction**
  The whole 4D volume (3 Spatial + 1 Temporal) is created from the DCE-MRI data.
- **BreastMask Extraction**[2]
  A binary mask representing only breast parenchyma is extracted
- **Preprocessing**
  Image processing techniques are applied with the aim of improving next stages
- **ROI Detection**
  Tumour lesions are identified as ROI (Region Of Interest), both malignant and benign
- **ROI Classification**
  Each ROI is classified according to its staging

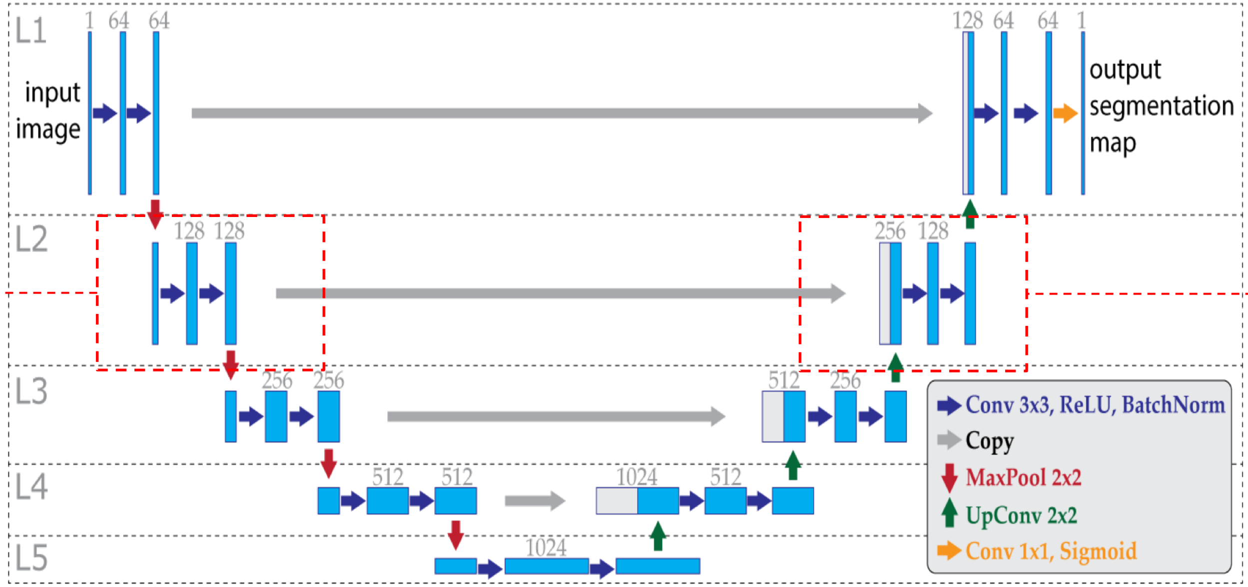
[1]Piantadosi, G., Marrone, S., Fusco, R., Sansone, M., Sansone, C.: A comprehensive computer-aided diagnosis for breast t1w dce-mri via quantitative dynamical features and spatio-temporal local binary patterns. IET Computer Vision (August 2018)

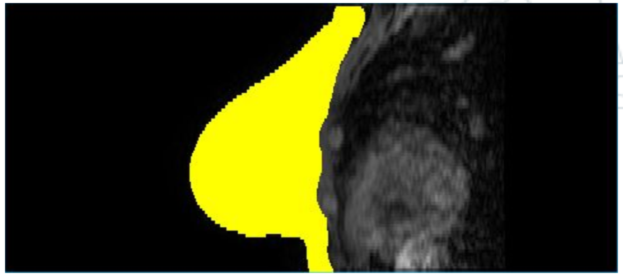
[2]Piantadosi, G., Sansone, M., Sansone, C.: Breast Segmentation in MRI via U-Net Deep Convolutional Neural Networks. In: 24rd International Conference on Pattern Recognition (ICPR) (2018)

Case Study: U-NET for Breast Tissues Segmentation

U-Shaped networks[3] consist of **two sides** and **several layers**. The left side performs a **contracting path** following the typical architecture of a convolutional neural network. On the other hand, the right side performs the **expansive path** with the same idea of the left side, but aiming to increase the image sizes.







Contracting path super-layer

1. 2D Conv.(3x3 kernel, zero-padding, stride 1x1)
2. Rectified Linear Unit (ReLU) activation
3. Batch Normalisation
4. 2D Conv. (3x3 kernel, zero-padding, stride 1x1)
5. Rectified Linear Unit (ReLU) activation
6. Batch Normalisation
7. Pooling (max-pool, stride 2x2)

Performs a down-sampling step halving the size of the image and doubling the size of the feature channels.

U-Net configuration in ICPR2018[2]: 5 main layers network, 23 convolution layers and 4 max-pooling layers, resulting in a total of 7.771.297 parameters.

[2] Piantadosi, G., Sansone, M., Sansone, C.: Breast Segmentation in MRI via U-Net Deep Convolutional Neural Networks. In: 24rd International Conference on Pattern Recognition (ICPR) (2018)

[3] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234/241. Springer (2015)

Expansive path super-layer

1. 2D Up-Conv. (2x2 kernel, zero-padding, stride 1x1)
2. Concatenation with feature-map from the same level of the contr. path.
3. 2D Conv. (3x3 kernel, zero-padding, stride 1x1)
4. Rectified Linear Unit (ReLU) activation
5. Batch Normalisation
6. 2D Conv. (3x3 kernel, zero-padding, stride 1x1)
7. Rectified Linear Unit (ReLU) activation
8. Batch Normalisation

The Up-Conv. performs the halving of the feature channels with a trainable kernel.

Evaluation Strategy

The network was trained by minimizing the task-specific loss: 1−DSC. Where DSC is the Dice Similarity Coefficient and n(·) represents the enclosed volume number of voxels and defined as:

$$DSC = (2 \cdot n(GS \setminus SEG))=(n(GS) + n(SEG))$$

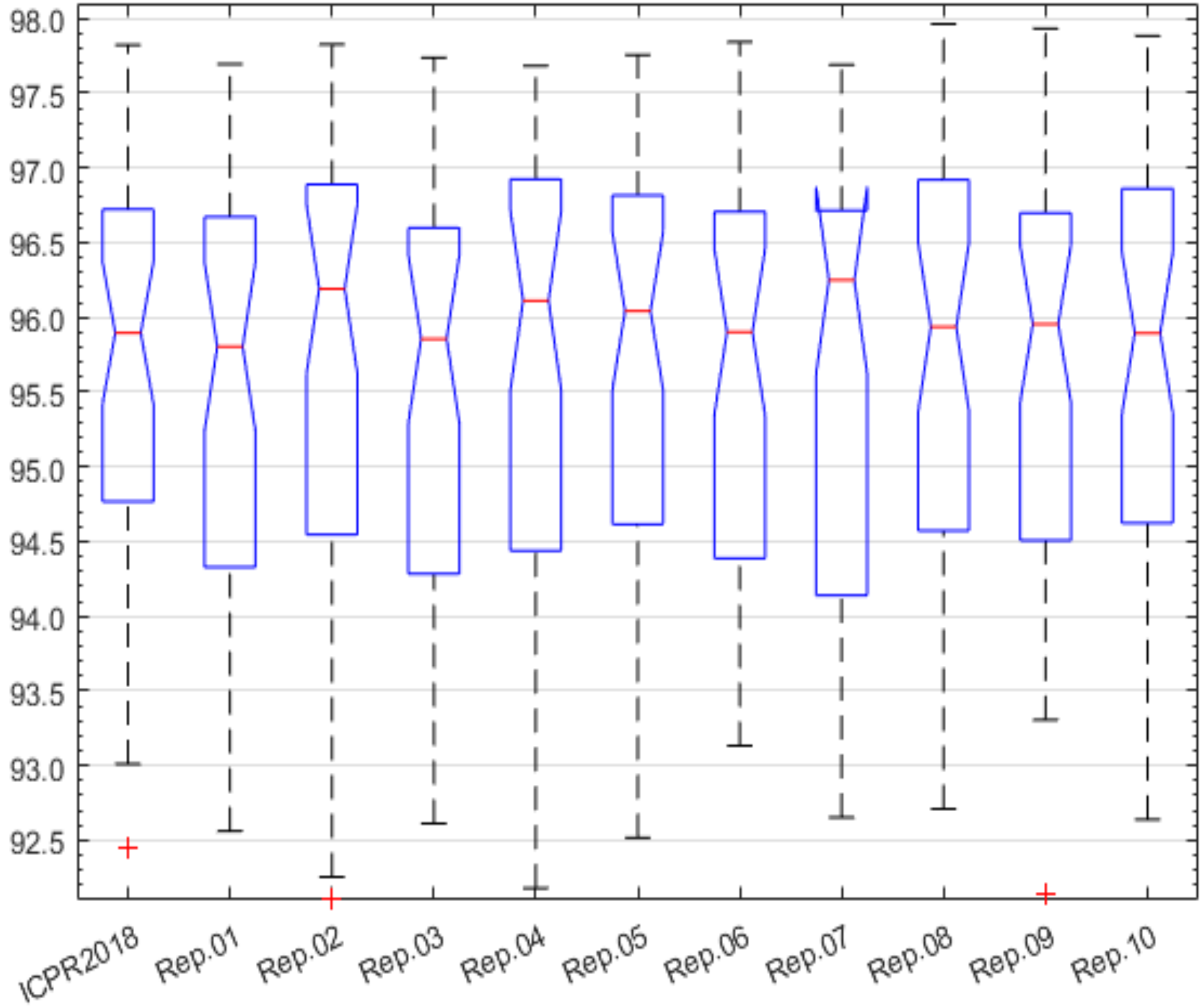
Software
Python 3.6
Keras (front-end)
TensorFlow 1.9 (back-end)

Hardware
Google Colaboratory (VM)
2 x Intel(R) Xeon(R) @ 2.2GHz CPUs
13GB RAM
NVIDIA K80 GPU (12GB GRAM)

- Patient-based 10-fold Cross Validation (CV),
- Repeatability evaluation: 50 execution with the same initialization seeds for the random numbers generators to try highlighting only the uncertainty due to random considerations introduced by the optimization tools' randomness.

Results

Repetitions	DSC [%]	LB [%]	UB [%]
ICPR2018[2]	**95.90**	**95.16**	**96.64**
Rep.01	95.80	95.24	96.37
Rep.02	96.19	95.62	96.75
Rep.03	95.85	95.38	96.39
Rep.04	96.11	95.69	96.57
Rep.05	96.04	95.15	96.62
Rep.06	95.90	95.02	96.60
Rep.07	96.25	95.29	96.52
Rep.08	95.93	95.44	96.56
Rep.09	95.95	95.38	96.36
Rep.10	95.89	95.35	96.43



Results obtained for the first 10 out of 50 Montecarlo executions of the 10-fold cross validation of our approach. The results presented in ICPR2018 are also reported in bold. Median values [left] with corresponding 95% confidence intervals (LB: LowerBound, UB: UpperBound) and boxplots [right] are reported .

2. *Piantadosi, G., Sansone, M., Sansone, C.: Breast Segmentation in MRI via U-Net Deep Convolutional Neural Networks. In: 24rd International Conference on Pattern*

Conclusions

Key Findings:

- We quantitatively highlighted the reproducibility problem of Convolutional Neural Networks (CNN) based approaches evaluating our deep learning approach for breast segmentation.
- We can state that our CNN-based model is stable to the different training executions since the confidence intervals obtained on the tests data overlap.

General Findings:

- This problem is not limited to the analyzed framework, Tenwsorflow, but lies in the NVIDIA CUDA Deep Neural Network (cuDNN) libraries.
- The randomness introduced with the advent of optimization engines for deep learning models, even if it may impact on the results of a reliable and reproducible research, only shift the attention on the statistical validity of the obtained outcomes.