Modified genetic algorithm for training of deep CNN architectures

Milan M. Čugurović, Faculty of Mathematics, UB

Nikola Dimitrijević, Microsoft Development Center Serbia

doc dr Stefan Mišković, Faculty of Mathematics, UB

Focus on

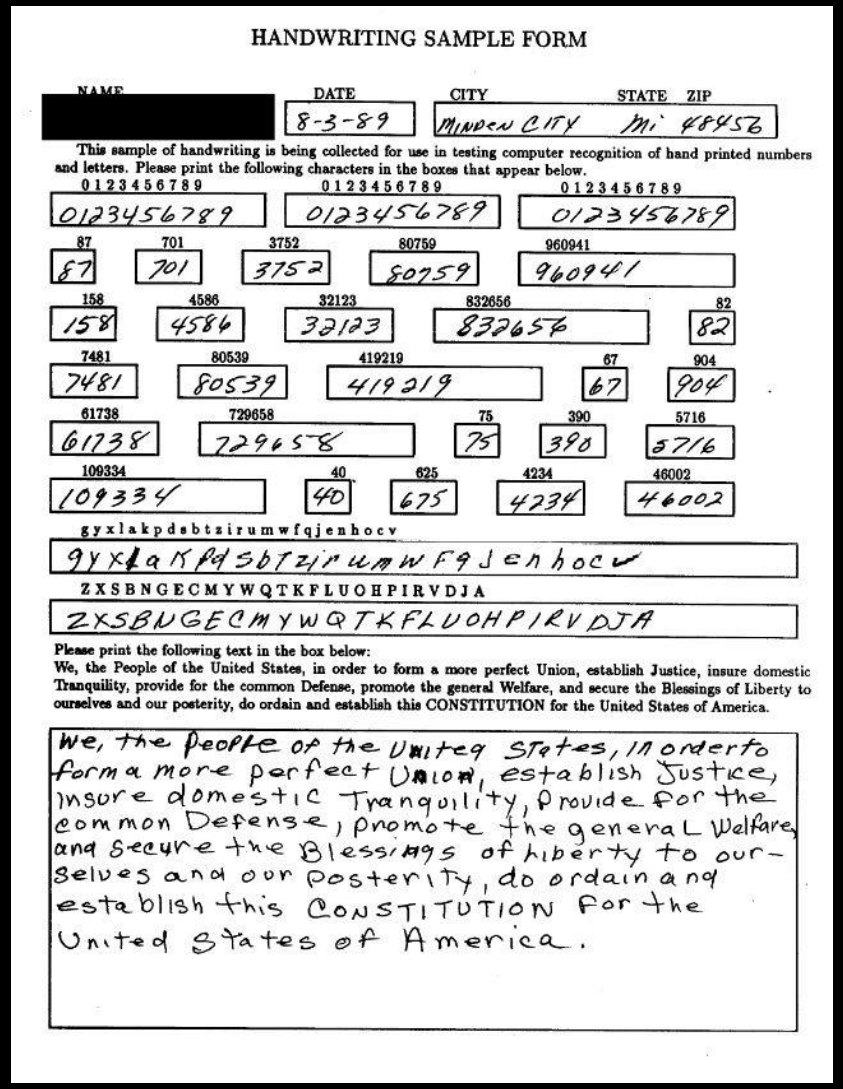
- Merge:
  - Deep Learning
  - Genetic Algorithms
  - Handwriting recognition
- Todays section: Mathematics and its applications today
  - Gradient, Partial Derivate, Function composition, etc.
- A kind of homage:
  - Petrović, Pejović, Dajović, Mateljević

MNIST dataset [LeCun et al., 1998]

- The most famous set of handwritten characters (digits)
- 60k+10k, 28x28x1
- Yann et. all, 1997
- Example:

NIST Special Database 19 [Grother, 1995]

- American National Institute of Standards and Technology
- >800k images, >3600 writers
- 128x128x1, 62 labels
- *EMNIST* [Cohen et al., 2017]



HANDWRITING SAMPLE FORM

NAME [redacted] DATE 8-3-89 CITY MINDEN CITY STATE MI ZIP 48456

This sample of handwriting is being collected for use in testing computer recognition of hand printed numbers and letters. Please print the following characters in the boxes that appear below.

0123456789 0123456789 0123456789

0123456789 0123456789 0123456789

87 701 3752 80759 960941

87 701 3752 80759 960941

158 4586 32123 832656 82

158 4586 32123 832656 82

7481 80539 419219 67 904

7481 80539 419219 67 904

61738 729658 75 390 5716

61738 729658 75 390 5716

109334 40 625 4234 46002

109334 40 675 4234 46002

g y x l a k p d s b t z i r u m w f q j e n h o c v

9 y x l a k p d s b t z i r u m w f q j e n h o c v

Z X S B N G E C M Y W Q T K F L U O H P I R V D J A

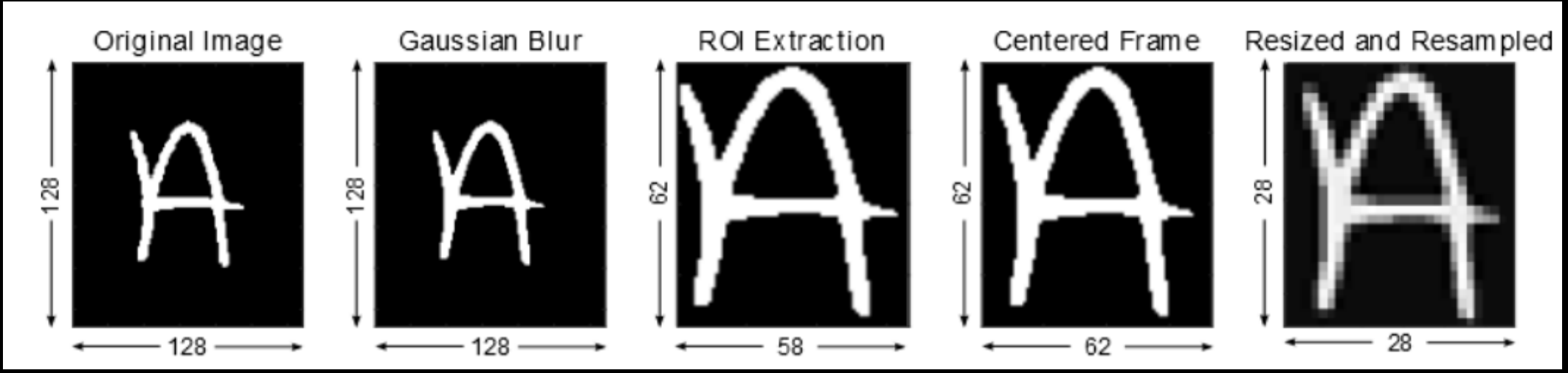
Z X S B N G E C M Y W Q T K F L U O H P I R V D J A

Please print the following text in the box below:
We, the People of the United States, in order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common Defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our posterity, do ordain and establish this CONSTITUTION for the United States of America.

We, the People of the United States, in order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common Defense, promote the general Welfare, and secure the Blessings of liberty to ourselves and our posterity, do ordain and establish this CONSTITUTION for the United States of America.

Doubled MNIST [*Čugurović et al.,2020*]

- Doubled in relation to MNIST in number of samples and in each sample resolution
- 120k+20k, 56x56x1
- 62 labels, ‘0’-’9’, ‘A’-’Z’, ‘a’-’z’
- [Cohen et al., 2017]:

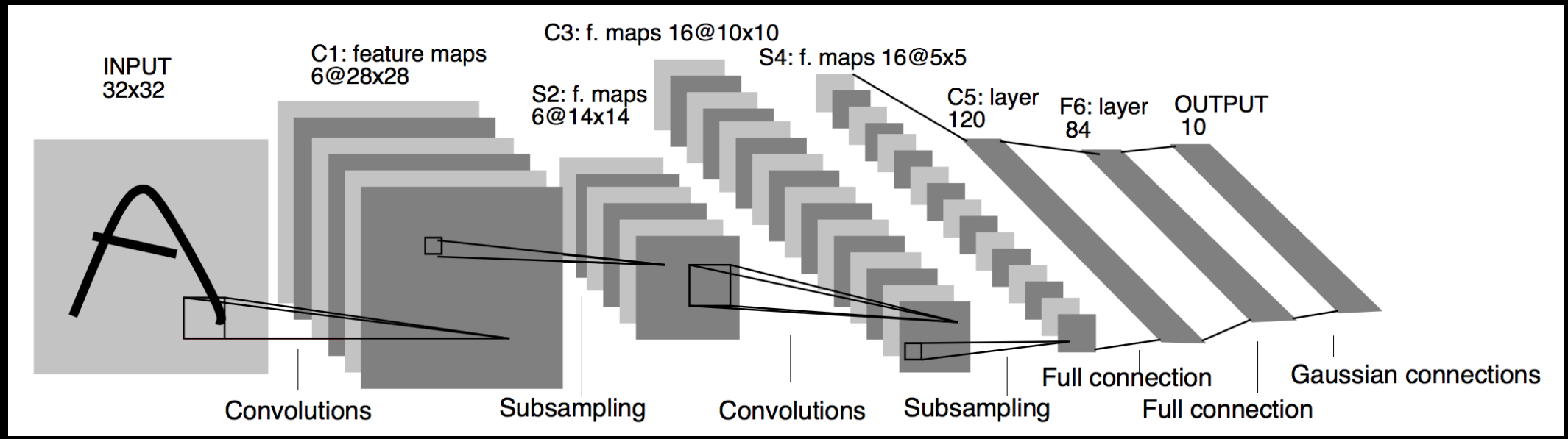


Doubled MNIST [*Čugurović et al.,2020*]

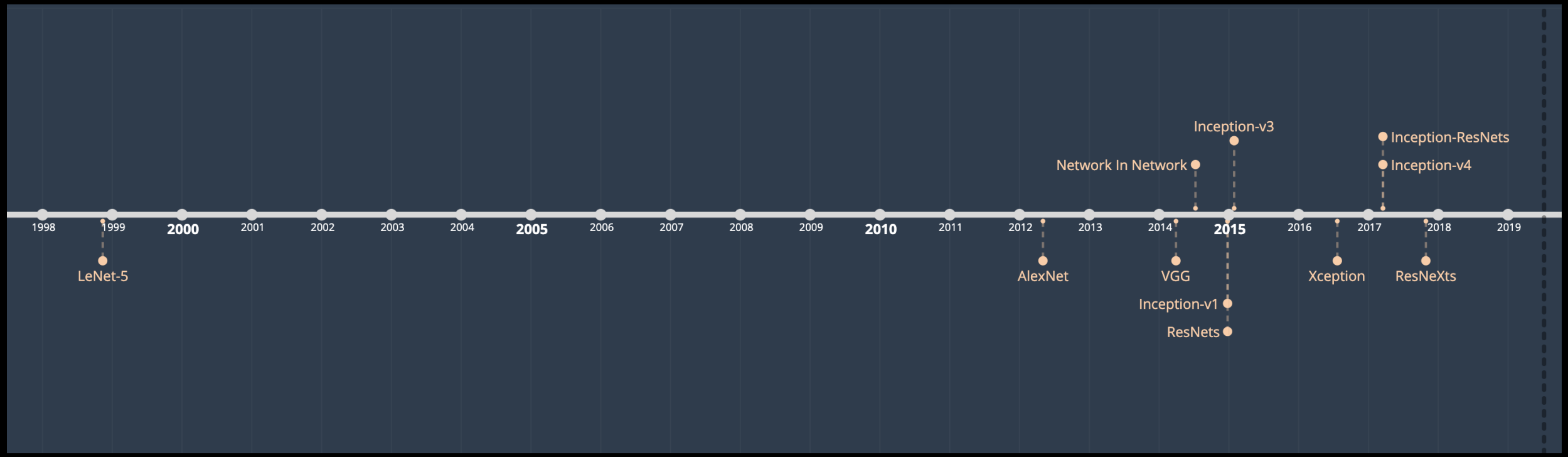
- Conversion process inspired by [Cohen et al., 2017]
- 128x128 pixel NIST binary images -> 56x56 pixel images with 8-bit grayscale resolution
  1. Each image loaded individually
  2. Fitting bounding box to character
  3. Extracting character (2 pixel padding)
  4. Apply Gaussian filter with standard deviation set to 2
  5. Centering into a square frame (preserving aspect ratio)
  6. Resizing image using bi-cubic interpolation
  7. Scaling pixels to 8-bit range

Genetic algorithm

- Inspired by paper Genetic CNN [Xie et al., 2017]
- How to design the best deep CNN architecture?
- Le-Net-5 [LeCun et al., 1998]

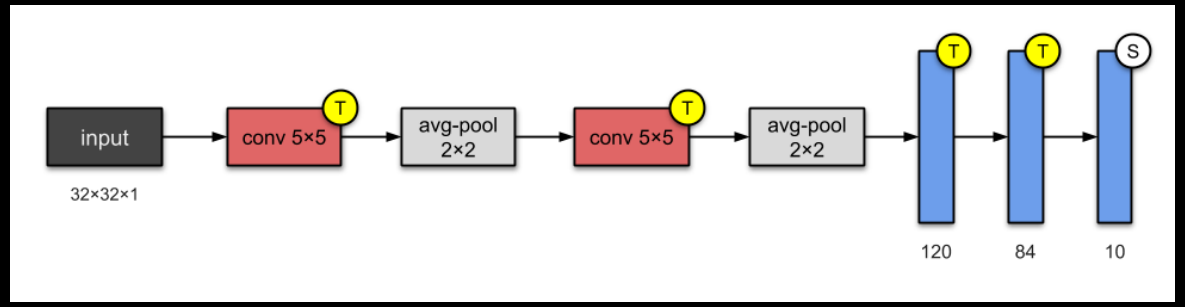


Genetic algorithm

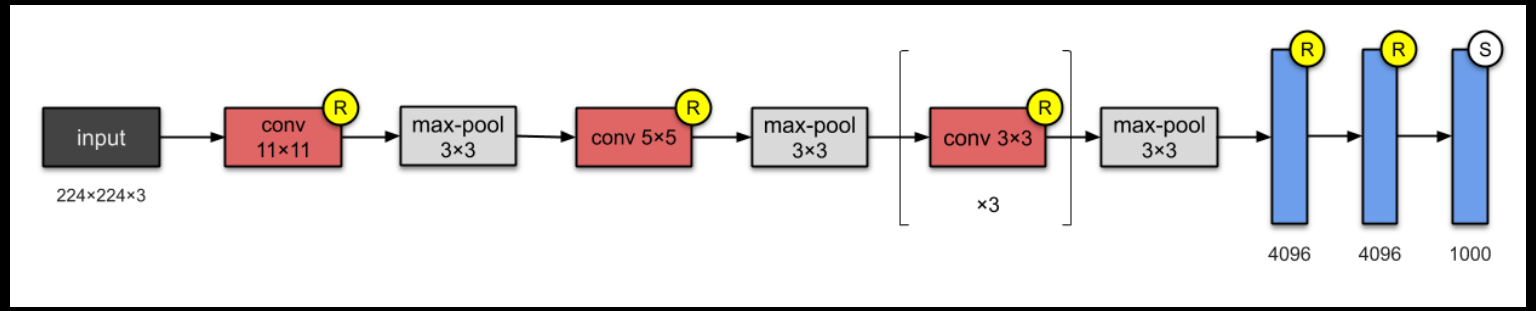


Architectures

- Lenet5 (1998, 60k params)



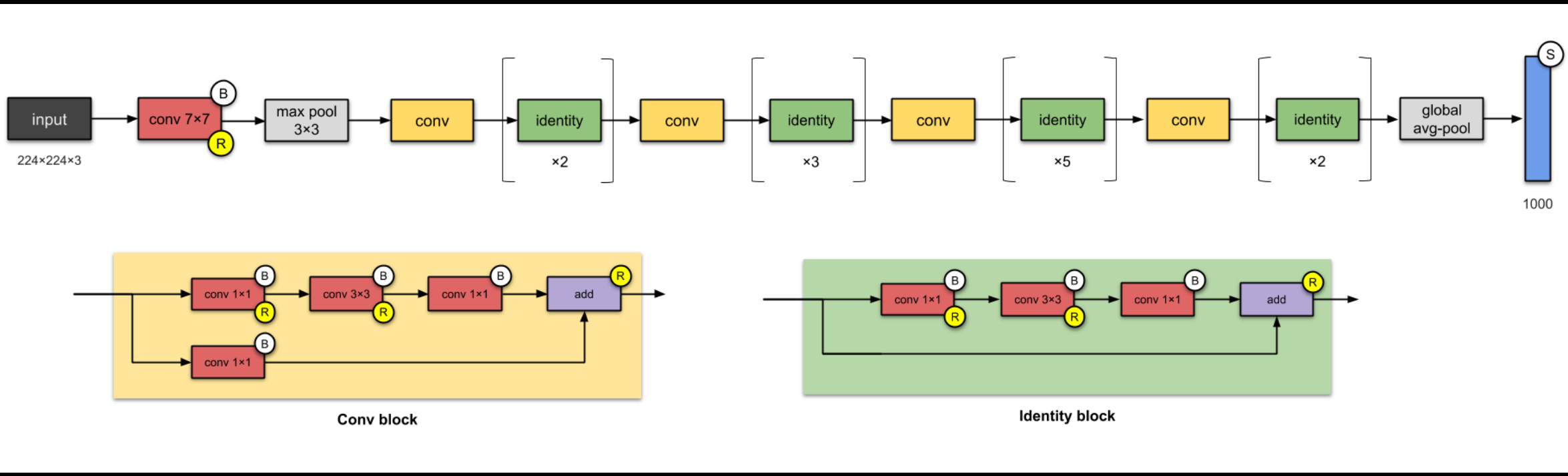
- AlexNet (2012, 60M params)



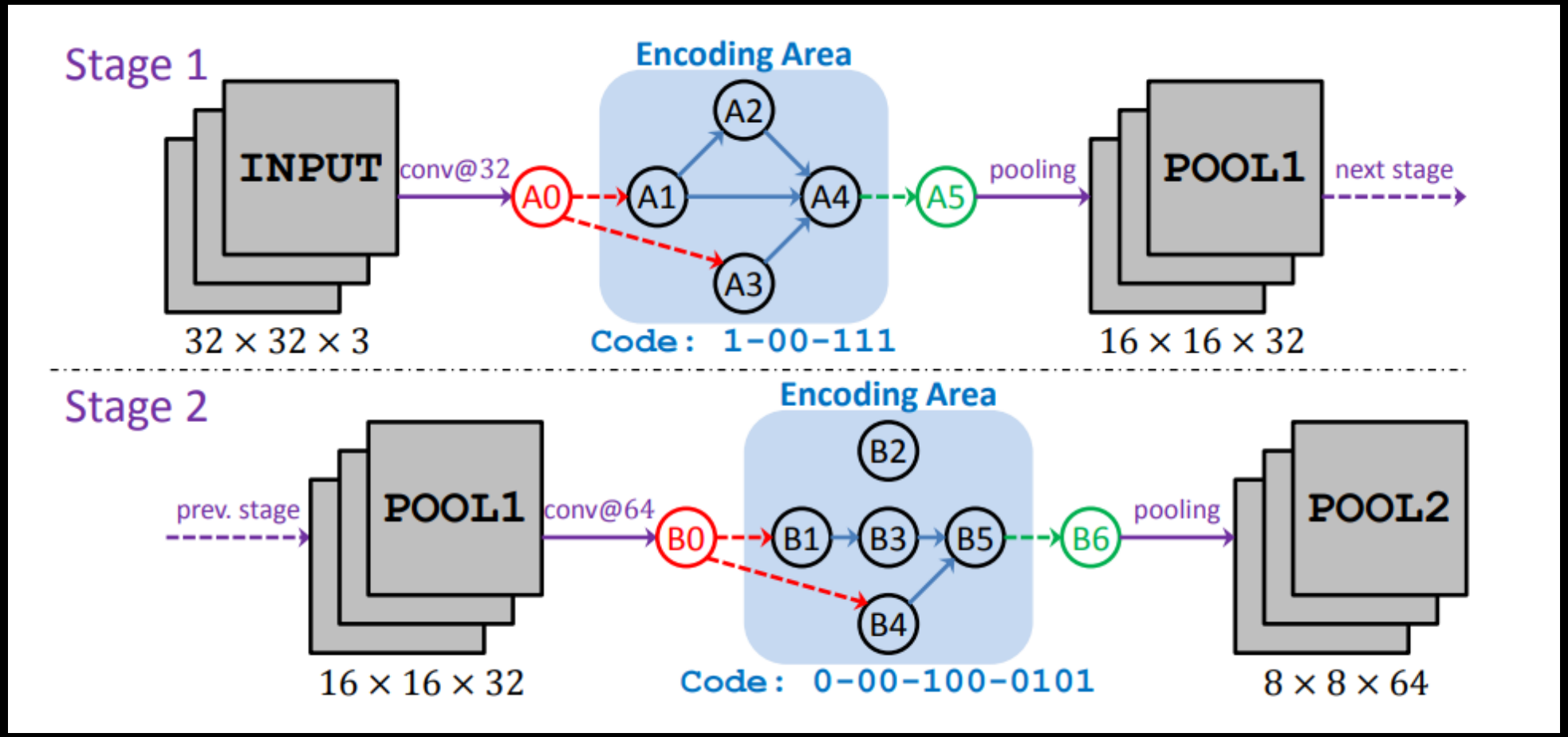
- VGG 16/19 (2014, 138M params)

Architectures

- ResNet (2015)
  - Skip connections
  - Very Deep CNN, up to 152 layers



Learn architecture [Xie et al., 2017]



GA [Xie et al., 2017]

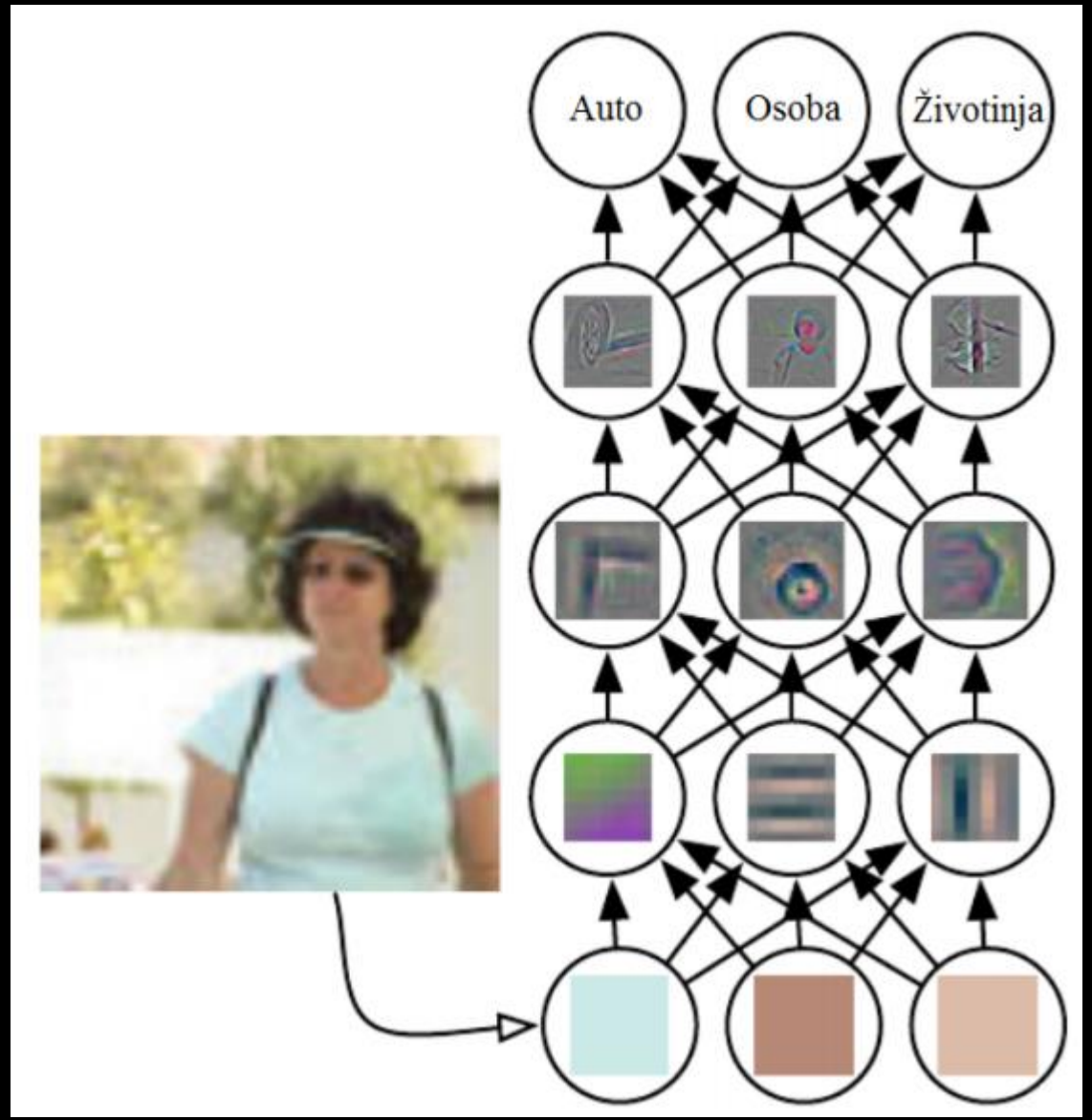
Algorithm 1 The Genetic Process for Network Design

1: **Input:** the reference dataset $\mathcal{D}$, the number of generations T, the number of individuals in each generation N, the mutation and crossover probabilities p_{M} and p_{C}, the mutation parameter q_{M}, and the crossover parameter q_{C}.
2: **Initialization:** generating a set of randomized individuals $\{\mathbb{M}_{0,n}\}_{n=1}^{N}$, and computing their recognition accuracies;
3: **for** $t = 1, 2, \ldots, T$ **do**
4: **Selection:** producing a new generation $\{\mathbb{M}'_{t,n}\}_{n=1}^{N}$ with a Russian roulette process on $\{\mathbb{M}_{t-1,n}\}_{n=1}^{N}$;
5: **Crossover:** for each pair $\{(\mathbb{M}_{t,2n-1}, \mathbb{M}_{t,2n})\}_{n=1}^{\lfloor N/2 \rfloor}$, performing crossover with probability p_{C} and parameter q_{C};
6: **Mutation:** for each non-crossover individual $\{\mathbb{M}_{t,n}\}_{n=1}^{N}$, doing mutation with probability p_{M} and parameter q_{M};
7: **Evaluation:** computing the recognition accuracy for each new individual $\{\mathbb{M}_{t,n}\}_{n=1}^{N}$;
8: **end for**
9: **Output:** a set of individuals in the final generation $\{\mathbb{M}_{T,n}\}_{n=1}^{N}$ with their recognition accuracies.

- time-consuming?

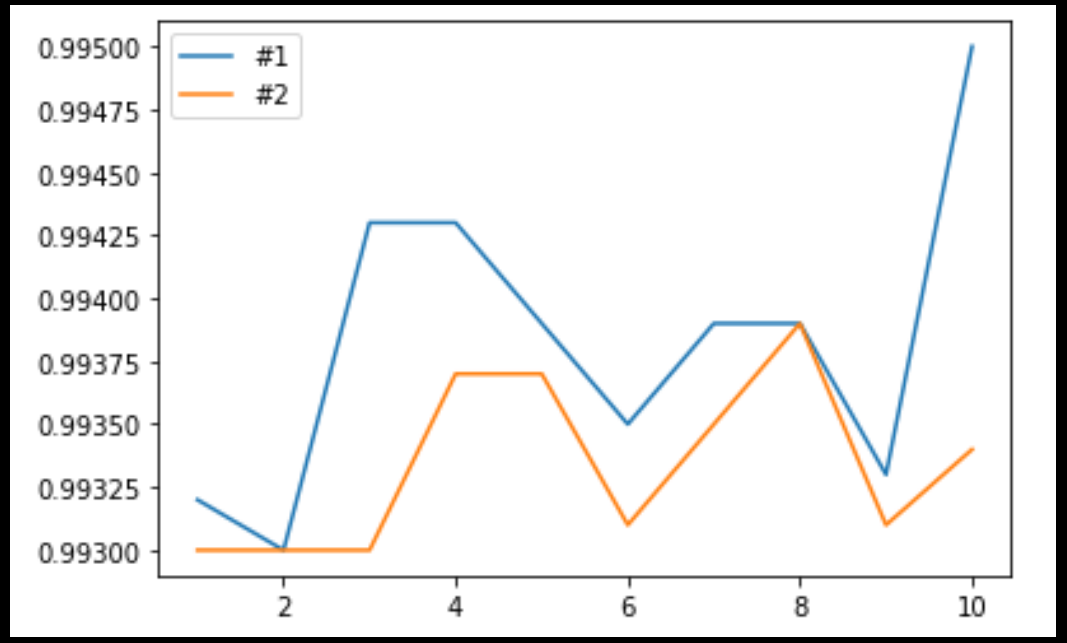
Our Contribution

- Reuse of learned weights (block level)
- Interpretability of CNN
- Fast adaptation to new attributes
- First 1, 2, 3?

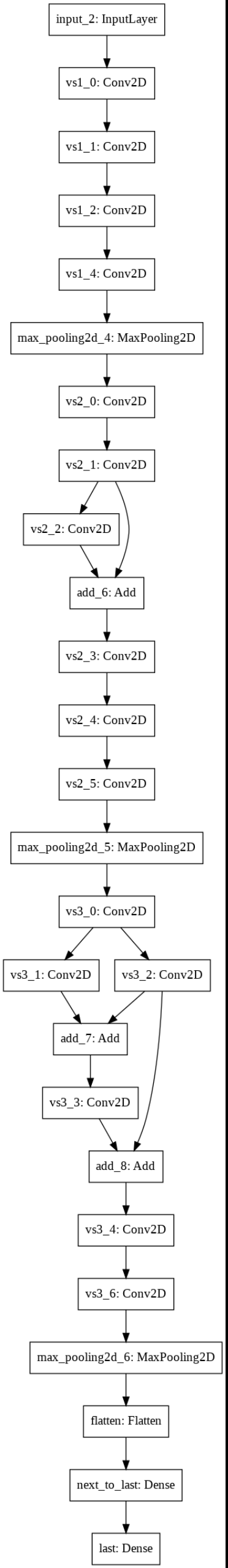


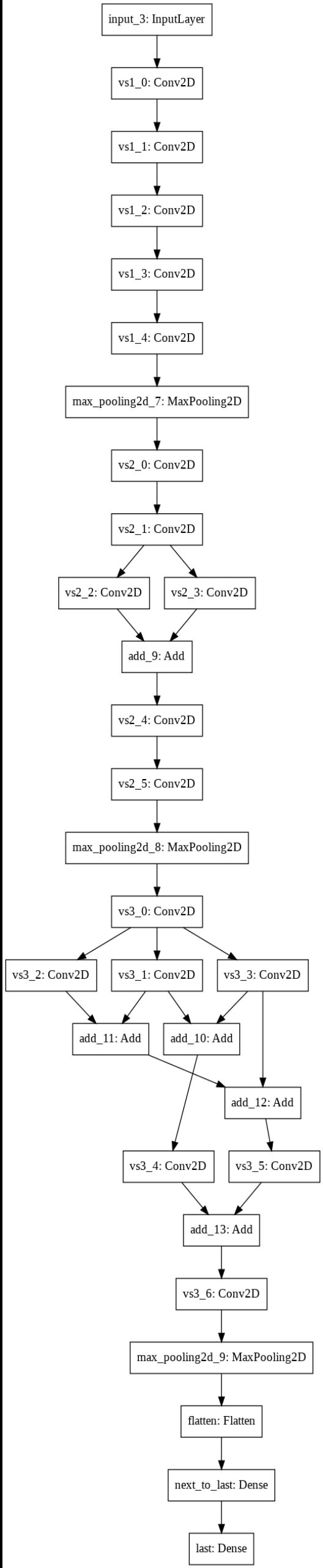
Our Results

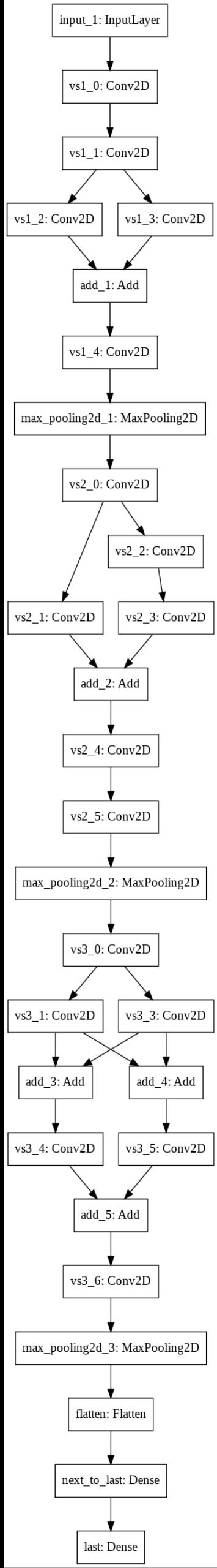
- Example 10x10x10
- Top2:



- In only 10 epochs, from 99.32 to 99.52 + Same as Biologicaly

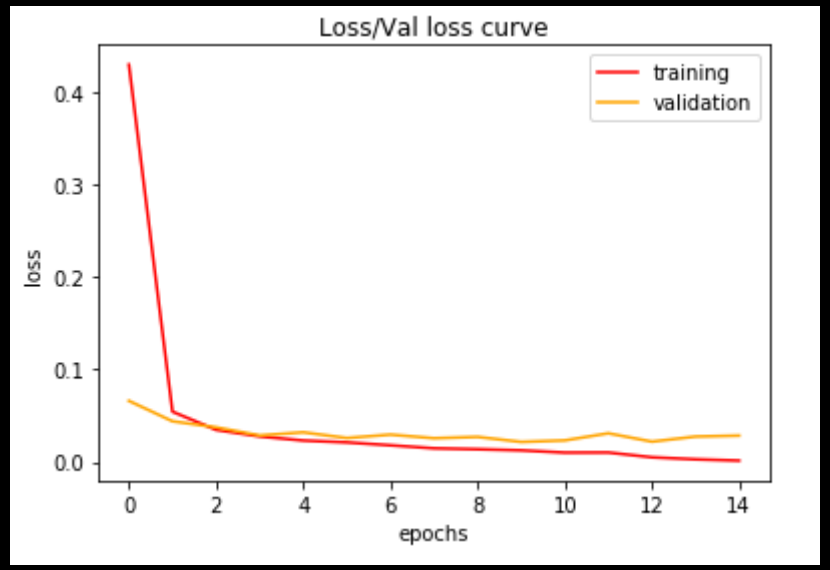

input_2: InputLayer
vs1_0: Conv2D
vs1_1: Conv2D
vs1_2: Conv2D
vs1_4: Conv2D
max_pooling2d_4: MaxPooling2D
vs2_0: Conv2D
vs2_1: Conv2D
vs2_2: Conv2D
add_6: Add
vs2_3: Conv2D
vs2_4: Conv2D
vs2_5: Conv2D
max_pooling2d_5: MaxPooling2D
vs3_0: Conv2D
vs3_1: Conv2D
vs3_2: Conv2D
add_7: Add
vs3_3: Conv2D
add_8: Add
vs3_4: Conv2D
vs3_6: Conv2D
max_pooling2d_6: MaxPooling2D
flatten: Flatten
next_to_last: Dense
last: Dense

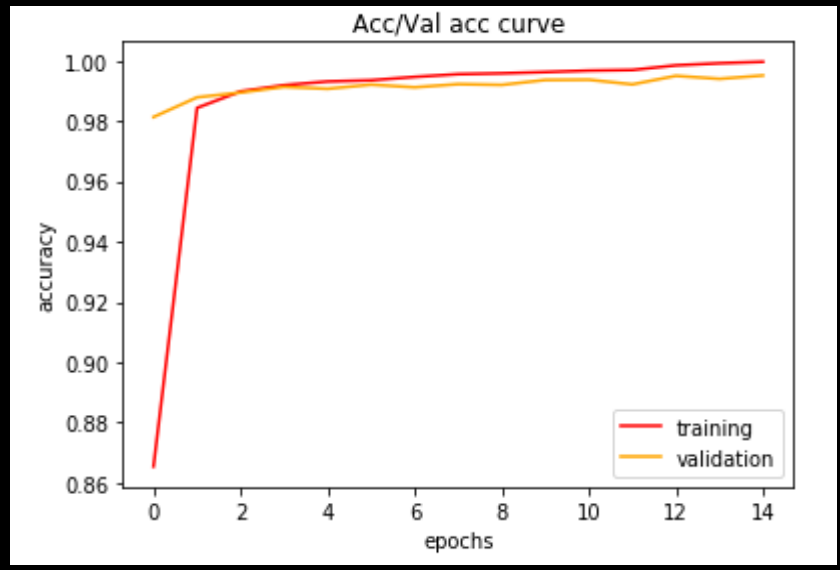

input_3: InputLayer
vs1_0: Conv2D
vs1_1: Conv2D
vs1_2: Conv2D
vs1_3: Conv2D
vs1_4: Conv2D
max_pooling2d_7: MaxPooling2D
vs2_0: Conv2D
vs2_1: Conv2D
vs2_2: Conv2D
vs2_3: Conv2D
add_9: Add
vs2_4: Conv2D
vs2_5: Conv2D
max_pooling2d_8: MaxPooling2D
vs3_0: Conv2D
vs3_2: Conv2D
vs3_1: Conv2D
vs3_3: Conv2D
add_11: Add
add_10: Add
add_12: Add
vs3_4: Conv2D
vs3_5: Conv2D
add_13: Add
vs3_6: Conv2D
max_pooling2d_9: MaxPooling2D
flatten: Flatten
next_to_last: Dense
last: Dense


input_1: InputLayer
vs1_0: Conv2D
vs1_1: Conv2D
vs1_2: Conv2D
vs1_3: Conv2D
add_1: Add
vs1_4: Conv2D
max_pooling2d_1: MaxPooling2D
vs2_0: Conv2D
vs2_2: Conv2D
vs2_1: Conv2D
vs2_3: Conv2D
add_2: Add
vs2_4: Conv2D
vs2_5: Conv2D
max_pooling2d_2: MaxPooling2D
vs3_0: Conv2D
vs3_1: Conv2D
vs3_3: Conv2D
add_3: Add
add_4: Add
vs3_4: Conv2D
vs3_5: Conv2D
add_5: Add
vs3_6: Conv2D
max_pooling2d_3: MaxPooling2D
flatten: Flatten
next_to_last: Dense
last: Dense

Time results

- MNIST:
  - 4x4 basic: 32min 44s, 4x4 modification: 18min 1s
  - 20x2 basic: 1h 22min 22s, 20x2 modification: 46min 55s
  - 2x20 basic: 25min 25s, 2x20 modification: 18min 18s
  - 20x20 basic: 2h 33min 54s, 20x20 modification: 1h 6min 25s [restricted /6]
- Doubled MNIST:
  - 29min 31s
  - Early stopping
  - 14 epochs:
  - 99.97/99.51/99.495





Thank you

- {milan_cugurovic, stefan}@matf.bg.ac.rs
- nikoladim@gmail.com

References

- [LeCun et al., 1998]
  - Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11):2278-2324, November 1998
- [Grother, 1995]
  - Grother, P. J. “NIST special database 19.”, *Handprinted forms and characters database, National Institute of Standards and Technology*, 1995
- [Cohen et al., 2017]
  - Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). “EMNIST: an extension of MNIST to handwritten letters.” *arXiv preprint arXiv:1702.05373*.
- [Čugurović et al., 2020]
  - Milan Čugurović, Nikola Dimitrijević, Stefan Mišković, “Modified genetic algorithm for training of deep CNN architectures”, Annaunced for 2020

References

- [Xie et al, 2017]
  - Xie, L., Yuille, A. Genetic cnn. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1379-1388), 2017.
- [Towards Data Science, towardsdatascience.com]
  - https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d#e4b1