AUTOMATED ARCHITECTURE-MODELING FOR CNNS

MANH KHOI DUONG

TABLE OF CONTENTS

1. Introduction
   1. Motivation
   2. Related Works
2. Methods
   1. Foundation: Bayesian Optimization
   2. Model finding
   3. Fine-tuning
3. Evaluation
   1. Dataset
   2. Final Results
   3. Comparison with related works
4. Conclusion and Future Works

INTRODUCTION

MOTIVATION

- Artificial Neural Networks (ANNs) are nowadays very popular and used widely in many tasks
- Among image classification datasets, Convolutional Neural Networks (CNNs) are used often
- But ANNs have **too many hyperparameters**!
- Not only the learning rate, dropout value etc. can be considered
  - Numerous ANNs architectures can be considered
  - What yields the best results?
- ANNs are still considered as a Black Box function
  - Unclear what makes the model learn best
  - Makes hyperparameter tuning even harder

TASK

- Use Bayesian Optimization [6] to find the optimal model architecture
- Final evaluation: Histological dataset of breast cancer

MOTIVATION: FIGURES

Fig. 1: Illustration of VGG16 by Kasthurirangan Gopalakrishnan

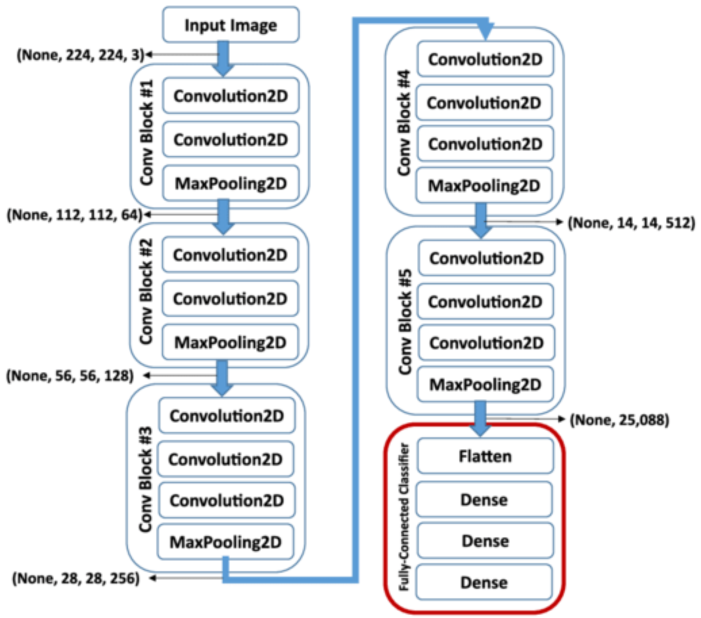
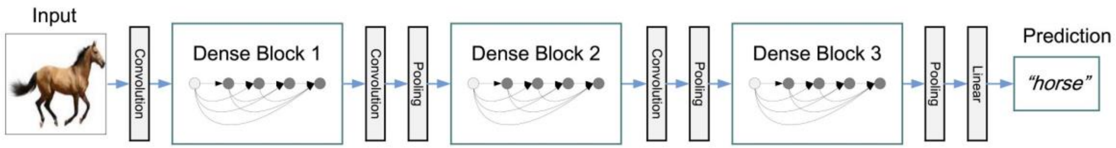


Fig. 2: Illustration taken from the DenseNet paper [1]



Many architectures can be proposed as seen in Figure 1,2

RELATED WORKS

- Grid Search, Random Search [2]
- Reinforcement learning methods
  - Neural architecture search with reinforcement learning (NAS) [3]
  - Progressive neural architecture search [4]
  - Learning transferable architectures for scalable image recognition [5]
  - Controller produces models and each model returns the validation accuracy
  - The validation acc. guides the controller to produce better models

FOUNDATION

FOUNDATION: BAYESIAN OPTIMIZATION [6]

- Bayesian Opt. is an algorithm for **global optimization**
  - The minimum or maximum of every function ***f*** can be found

BAYESIAN OPTIMIZATION: INTUITIVE IDEA

- Suppose function *f* has to be **minimized**
- It is possible to evaluate a certain point of *f*
- ➔ For every x_i we get a $y_i = f(x_i)$ for all $i \in \mathbb{N}$

BAYESIAN OPTIMIZATION: INTUITIVE IDEA

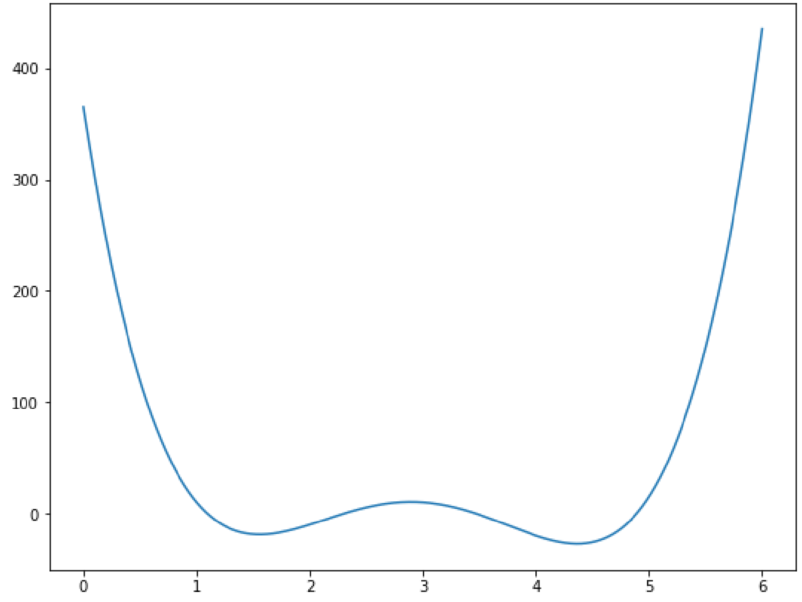
- Assume that function *f* is very expensive to calculate
- **Aim: min(*f*) by evaluating as few points as possible**
- **Bayesian inference: Use information given by the regression model to find the minimum**
  - **Regions which where not discovered yet should be sampled to escape the local minimum**

BAYESIAN OPTIMIZATION: INTUITIVE IDEA

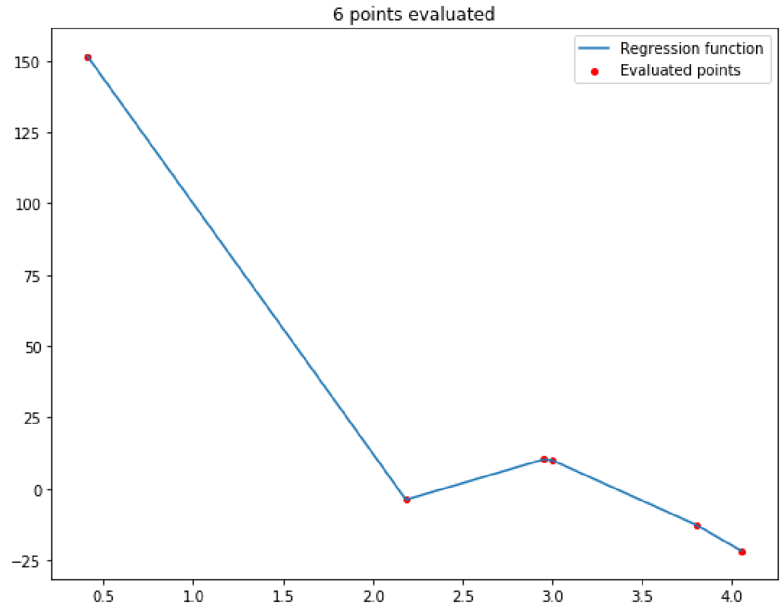
1. Create a regression model p based on the observed points
2. Sample next point:
   - **Explore** regions with higher uncertainty or **exploit** where the minimum could be found
   - Exploration-Exploitation trade-off
   - Acquisition function tells which point has to be evaluated next
3. Repeat 1.-2.

BAYESIAN OPTIMIZATION WITH GP'S

Real black box function f (unknown)

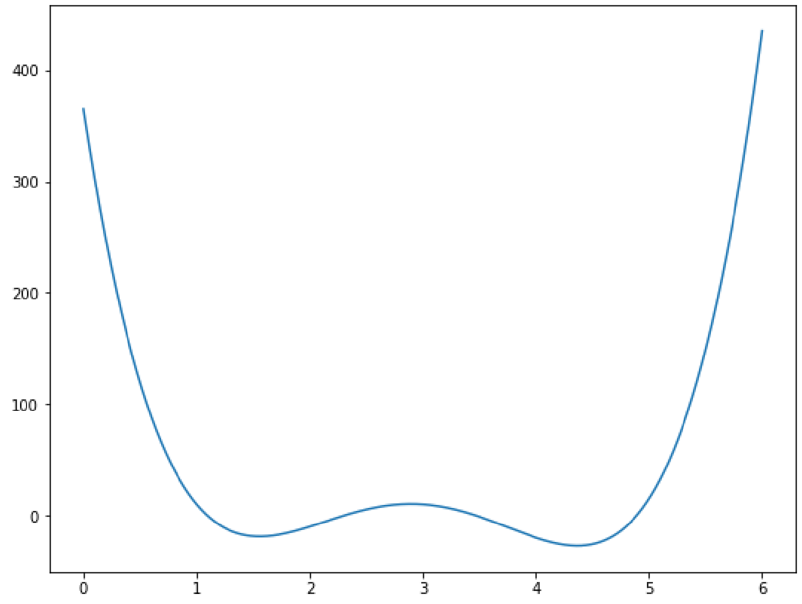


Approximation model p

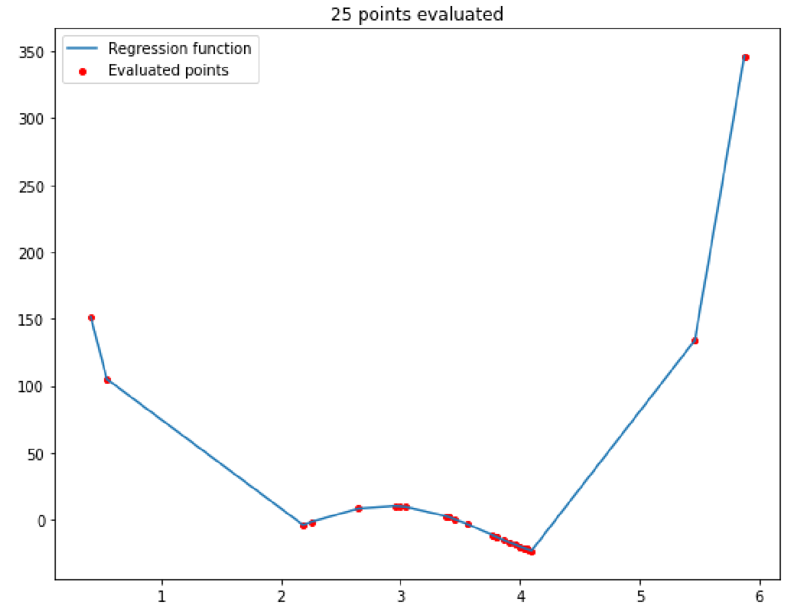


BAYESIAN OPTIMIZATION WITH GP'S

Real black box function f (unknown)

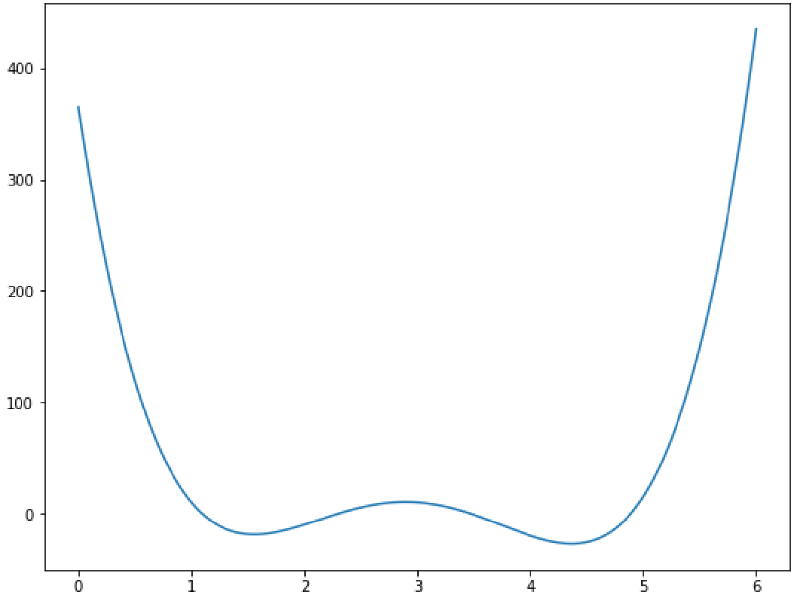


Approximation model p

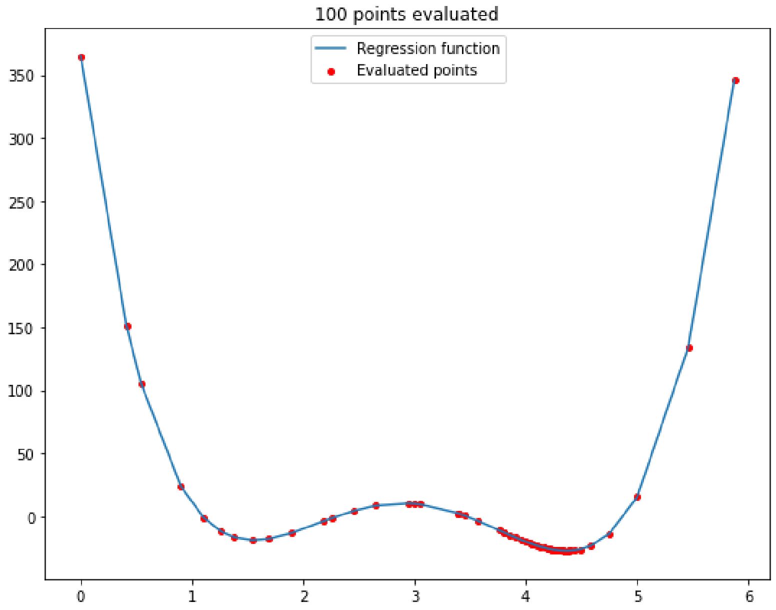


BAYESIAN OPTIMIZATION WITH GP'S

Real black box function f (unknown)



Approximation model p



METHODS

MODEL FINDING

Definition

- Let function f be a neural network which takes in the hyperparameter vector $\vec{x}$, training data D, validation data D_{val}, training label y, validation label y_{val}.
- Let f return the validation accuracy. Due to the minimization of the objective, the validation accuracy should be returned negated.
- Let $\tilde{f}$ be f where the validation accuracy is returned negatively.
- The objective is: $argmin_{\vec{x}}\ \tilde{f}(\vec{x}, D, y, D_{val}, y_{val})$

Bayesian Opt. applied on Hyperparam.opt.

Algorithm 1 Bayesian Optimization

```
procedure BAYESOPT(x⃗, D, y, D_val, y_val)
    for i ← 1, n do
        x⃗_next ← argmax_x⃗ acq(x⃗, D̃)                    ▷ acq is the acquisition function
        val_accuracy ← f̃(x⃗_next, D, y, D_val, y_val)
        if val_accuracy < best_accuracy then
            best_accuracy ← val_accuracy
            x⃗_best ← x⃗_next
        end if
        D̃ ← D̃ ∪ (x⃗_next, val_accuracy)
        Update Gaussian process regression based on D̃
    end for
    return x⃗_best
end procedure
```

MODEL FINDING: ARCHITECTURE

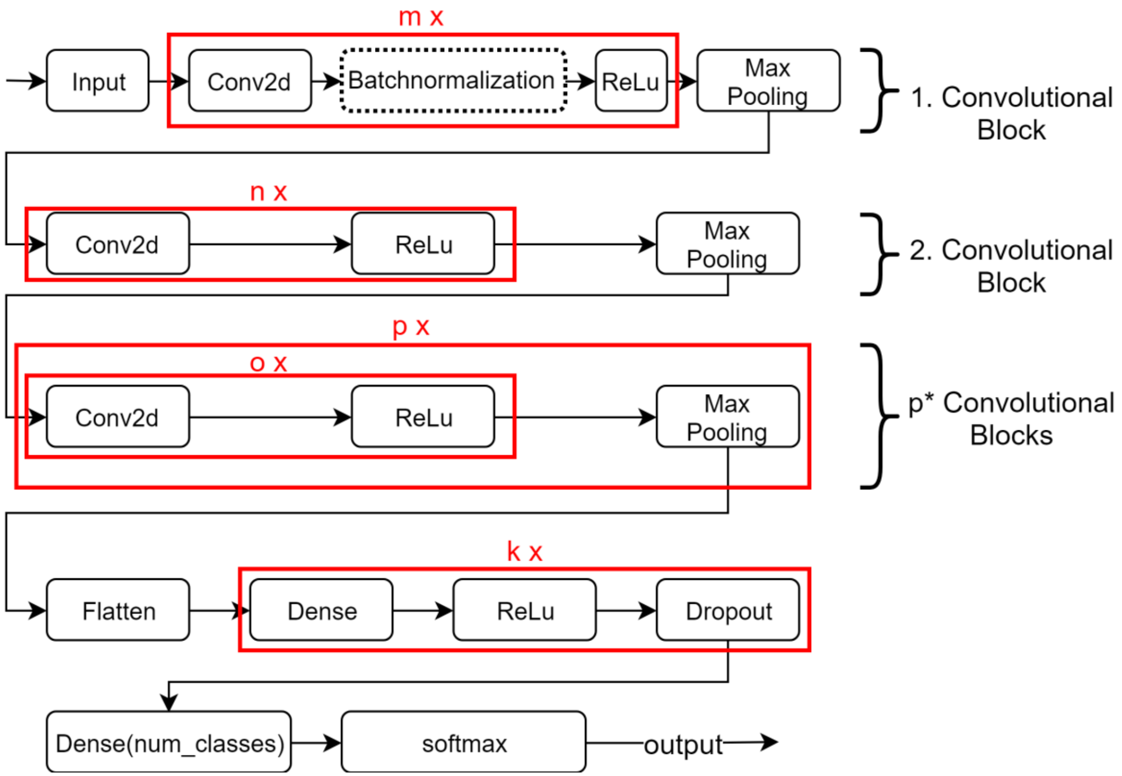


Fig. 4: Possible model architectures. m, n, o, p, k are hyperparameters

FINE-TUNING

- Same algorithm, $\tilde{f}$ is modifiable in other parts
- $\tilde{f}$ just takes a different hyperparameter vector $\vec{x}$
- Arbitrary choice of hyperparameters
  - Choice of hyperparameters can be chosen differently
  - Search space can be defined arbitrary

FINE-TUNING: DETAILS

- Eight hyperparameters were optimized on:
  - Output dimension of the Conv Blocks (3)
  - Kernel size of every Conv filter
  - Learning rate, dropout, number of dense nodes
  - L2-regularization

EVALUATION

DATASET [7]

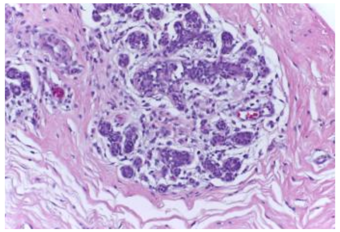


Fig. 5: Normal

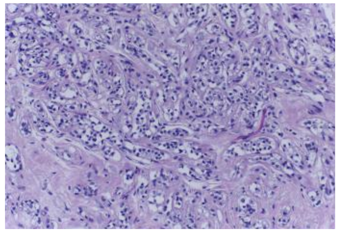


Fig. 6: Benign

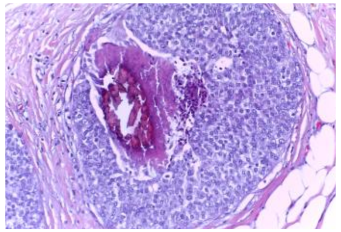


Fig. 7: In Situ

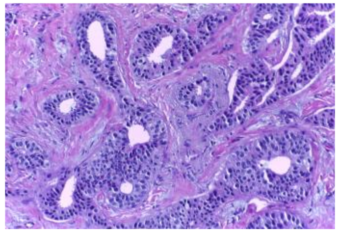


Fig. 8: Invasive

- Breast Cancer Dataset
- Histological images: tissue extracted and colorized with haematoxylin and eosin
- Magnification: 200x

DATASET [7]

- Information
  - Dataset consists of 249 training data and 20 test data
  - Dataset is mostly equally distributed
- Task
  - Four-class classification (benign, normal, in situ, invasive)
  - Two-class classification (noncarcinoma, carcinoma)
    - Benign, normal and summarized into noncarcinoma
    - In Situ and invasive are summarized into carcinoma
- Evaluation
  - In this work, the test set has been used as the validation data
  - For comparable results: only the validation accuracy is compared against similar works

DATA AUGMENTATION

- Augmentations:
  - Flipping
  - Mirror
  - Rotation
  - Random contrast changes
- Training data increased by factor 8 ➔ 249 * 8 = 1992 samples

EXPERIMENTATION SETTINGS

- Model finding
  - Create 600 models
  - Each model is trained for 40 epochs
- Fine-tuning
  - 1000 hyperparameters are tested
  - Best model is trained on those hyperparameters
  - Each hyperparameter setting is trained for 60 epochs
- Best model found (with fine-tuning) is trained further for 10000
  - Further training can decrease the validation acc due to overfitting:
  - Make use of early stopping

INSPECTING BAYESIAN OPTIMIZATION

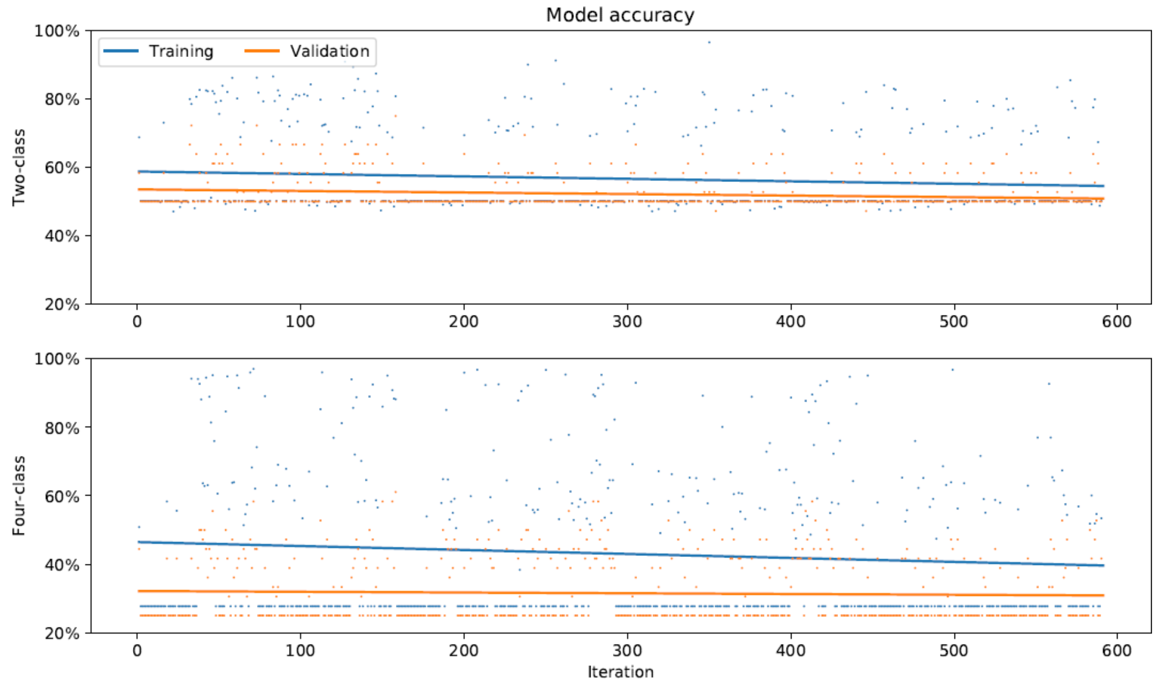


Fig. 9: Accuracy of the created models on both classification tasks. Linear Regression describes the course of the points.

- Points appear to be very random
- For both tasks, lots of models could not learn
  - Accuracy as good as the threshold
  - Randomly bad weight initialization
  - Maybe hard task (?)
  - High exploration rate
  - Sign that neural networks are chaotic systems

FINAL RESULTS

- Best models fine-tuned and trained further
- End result:
  - 69% accuracy for four classes
  - 89% accuracy for two classes

COMPARISON WITH RELATED WORKS

Team	4-class acc.	2-class acc.	Approach
Aditya et al. [8]	85%	93%	Transfer learning: Inception-v3 [11]
Kamyar et al. [9]	95%	-	Transfer learning: Inception-v3 [11]
Wajahat et al. [10]	81%	-	Transfer learning: AlexNet [12]
This work	**69%**	**89%**	**Automated Architecture-Modeling**

- [8], [9], [10] use a more extensive dataset (400 training samples) provided by the ICIAR Grand Challenge [13]
- Only participants had access to this dataset
- This work comes close with [8] when comparing 2-class results
- Inferior in the 4-class task

RESULTS

- Results not completely comparable
  - Dataset used in this work was more limited
  - Result also depends on the preprocessing part
    - Aditya et al. [8] and Kamyar et al. [9] use similar techniques
    - Still different results
  - Validation set different
  - Test accuracy not comparable for the same reason

CONCLUSION AND FUTURE WORKS

- Conclusion
  - Results good in the binary classification task
  - Results inferior in the four-class classification task
    - Lack of experimentation
    - Results not comparable
- Future Works
  - Allow the creation of nonsequential, recurrent models etc.
  - Search algorithm to find the optimal model
    - Lots of models do not learn from the data
    - Limit the training process
    - Number of epochs: Depth of tree
    - Set of models: Breadth of tree

THANK YOU FOR LISTENING

REFERENCES

- [1] Densely Connected Convolutional Networks
  - Gao Huang, Zhuang Liu and Kilian Q. Weinberger, arXiv preprint arXiv:1611.01578 (2016)
- [2] Random Search for Hyperparameter Optimization
  - James Bergstra and Yoshua Bengio, Journal of Machine Learning Research (2012)
- [3] Neural architecture search with reinforcement learning
  - Zoph, Barret and Quoc V. Le, arXiv preprint arXiv:1611.01578 (2016)
- [4] Progressive neural architecture search
  - Liu, Chenxi, et al., arXiv preprint arXiv:1712.00559 (2017)
- [5] Learning transferable architectures for scalable image recognition
  - Zoph, Barret, et al., CVPR (2018)
- [6] Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization
  - Jones, Donald R.; Schonlau, Matthias; Welch, William J., Journal of Global Optimization (1998)
- [7] Classification of breast cancer histology images using Convolutional Neural Networks
  - Teresa Araújo et al., Public Library of Science (PLOS) (2017)

REFERENCES

- [8] Classification of Breast Cancer Histology using Deep Learning
  - Golatkar, Aditya; Anand, Deepak; Sethi, Amit, CoRR, abs/1802.08080 (2018)
- [9] Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification
  - Nazeri, Kamyar; Aminpour, Azad; Ebrahimi, Mehran, International Conference Image Analysis and Recognition. Springer (2018)
- [10] Classification Of Breast Cancer Histology Images Using ALEXNET
  - Nawaz, Wajahat; Ahmed, Sagheer; Tahir, Muhammad; Khan, Hassan (2018)
- [11] Rethinking the Inception Architecture for Computer Vision
  - Szegedy, Christian; Vanhoucke, Vincent; Ioffe, Sergey; Shlens, Jonathon; Wojna, Zbigniew, CoRR, abs/1512.00567 (2015)
- [12] ImageNet Classification with Deep Convolutional Neural Networks
  - Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E, Advances in Neural Information Processing Systems 25 (2012)

REFERENCES

- [13] BACH: Grand Challenge on Breast Cancer Histology Images
  - Aresta, Guilherme; Araújo, Teresa; Kwok, Scotty; Chennamsetty, Sai Saketh; P., Mohammed Safwan K.; Varghese, Alex; Marami, Bahram; Prastawa, Marcel; Chan, Monica; Donovan, Michael J.; Fernandez, Gerardo; Zeineh, Jack; Kohl, Matthias; Walz, Christoph; Ludwig, Florian; Braunewell, Stefan; Baust, Maximilian; Vu, Quoc Dang; To, Minh Nguyen Nhat; Kim, Eal; Kwak, Jin Tae; Galal, Sameh; Sanchez-Freire, Veronica; Brancati, Nadia; Frucci, Maria; Riccio, Daniel; Wang, Yaqi; Sun, Lingling; Ma, Kaiqiang; Fang, Jiannan; Koné, Ismaël; Boulmane, Lahsen; Campilho, Aurélio; Eloy, Catarina; Polónia, António; Aguiar, Paulo
  - CoRR, abs/1808.04277 (2018)