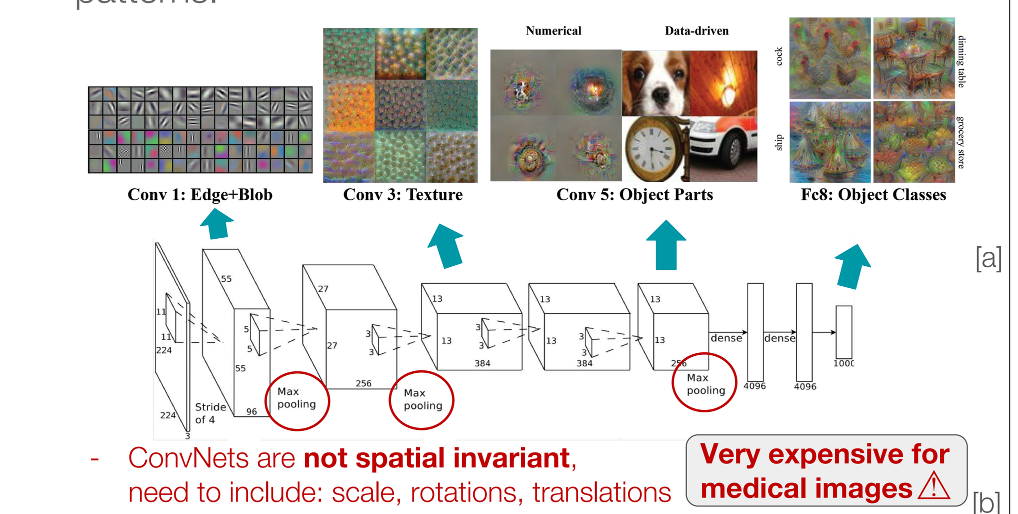
Capsule Networks against Medical Imaging Data Challenges

Amelia Jiménez-Sánchez[1], Shadi Albarqouni[2], Diana Mateus[3]

[1]BCN MedTech, DTIC, Universitat Pompeu Fabra, Spain
[2]Computer Aided Medical Procedures, Technische Universität München, Germany
[3]Laboratoire des Sciences du Numérique de Nantes, UMR 6004, Centrale Nantes, France

INTRODUCTION

- The ability of Convolutional Networks (**ConvNets**) to extract **meaningful and hierarchical feature representations** allow them to encode complex patterns.

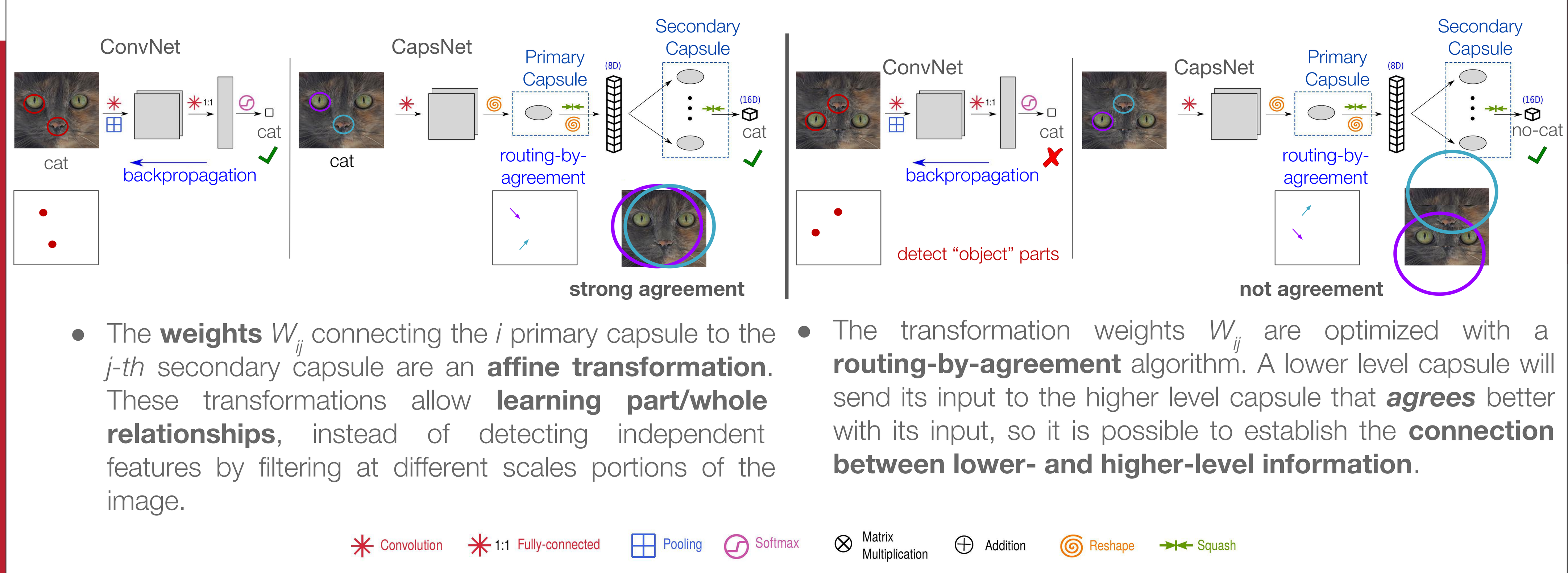


- They require **large amounts of annotated data** to represent the full variation of the classes.

[a] http://vision03.csail.mit.edu/cnn_art/index.html [b] https://www.flickr.com/ #cat

CAPSULE NETWORKS

- Capsule Networks (**CapsNets**) were recently introduced to cope with spatial invariance[1]. They are designed to **learn** the **pose** of the class instance together with its **presence**.
- Therefore, less variations of the instance are required, i.e. **fewer images**. This brought our attention to **medical datasets**, because they are frequently **small** and **highly imbalanced**.

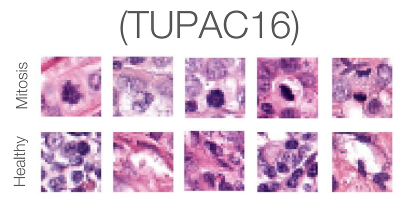


- The **weights** W_{ij} connecting the i primary capsule to the j-th secondary capsule are an **affine transformation**. These transformations allow **learning part/whole relationships**, instead of detecting independent features by filtering at different scales portions of the image.
- The transformation weights W_{ij} are optimized with a **routing-by-agreement** algorithm. A lower level capsule will send its input to the higher level capsule that ***agrees*** better with its input, so it is possible to establish the **connection between lower- and higher-level information**.

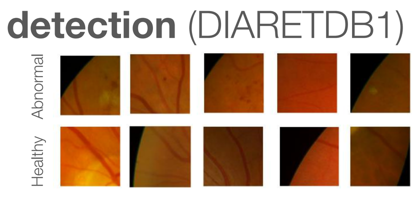
DATASETS

Our experiments are evaluated on **4 publicly available** datasets for two vision (MNIST and Fashion-MNIST) and two medical (TUPAC16 and DIARETDB1) datasets.

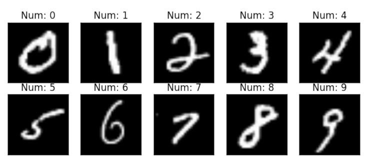
i) **Mitosis detection** (TUPAC16)



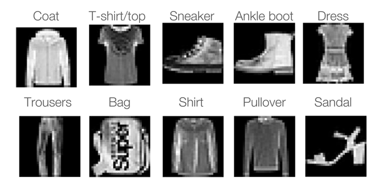
ii) **Diabetic retinopathy detection** (DIARETDB1)



iii) Handwritten Digit Recognition (MNIST)



iv) Clothes Classification (Fashion-MNIST)



ARCHITECTURES

	Conv1	Pool1	Conv2	Pool2	Conv3	-	FC1	Drop	FC2	#Params.
LeNet	5 × 5 6 ch	2 × 2	5 × 5 16 ch	2 × 2	✗	-	1 × 1 120 ch	✗	1 × 1 84 ch	60K
Baseline	5 × 5 256 ch	✗	5 × 5 256 ch	✗	5 × 5 128 ch	-	1 × 1 328 ch	✓	1 × 1 192 ch	35.4M
	Conv1	**Pool1**	**Conv2**	**Pool2**	**Caps1**	**Caps2**	**FC1**	**Drop**	**FC2**	**#Params.**
CapsNet	9 × 9 256 ch	✗	9 × 9 256 ch	✗	1152 caps 8D	N_k caps 16D	1 × 1 512 ch	✗	1 × 1 1024 ch	8.2M

Table 1. Details of each of the architectures. For convolution, we specify the size of the kernel and the number of output channels. In the case of pooling, the size of the kernel. And for capsule layers, first, the number of capsules and, in the second row, the number of dimensions of each capsule.

RESULTS

Hypothesis: We argue that CapsNet will perform better than ConvNets under medical data challenges.

(1) How do networks behave under decreasing **amounts of training data**?

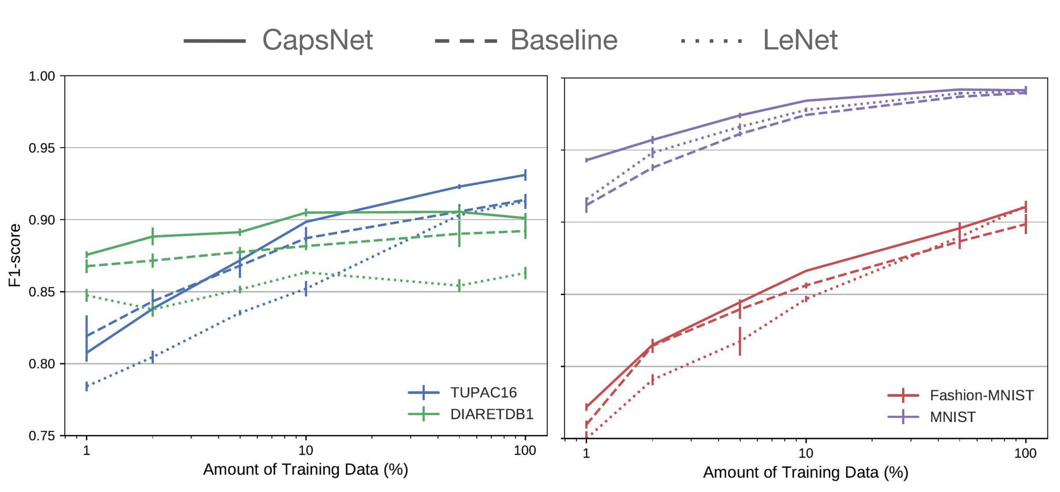


Figure 1. Mean F_1-score using different amounts of training data.

(2) Is there a change in their response to **class-imbalance**?

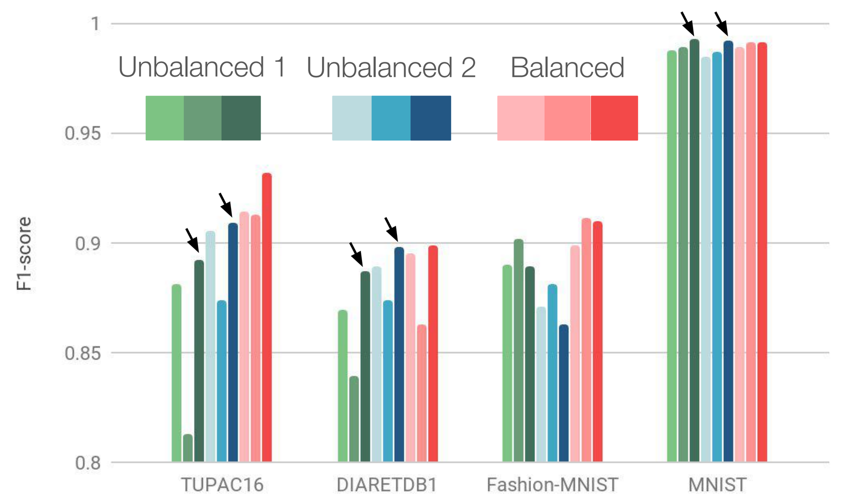


Figure 2. Mean F_1-score reported for different class-imbalance scenarios.

(3) Is there any benefit from **data augmentation** as a complementary strategy?

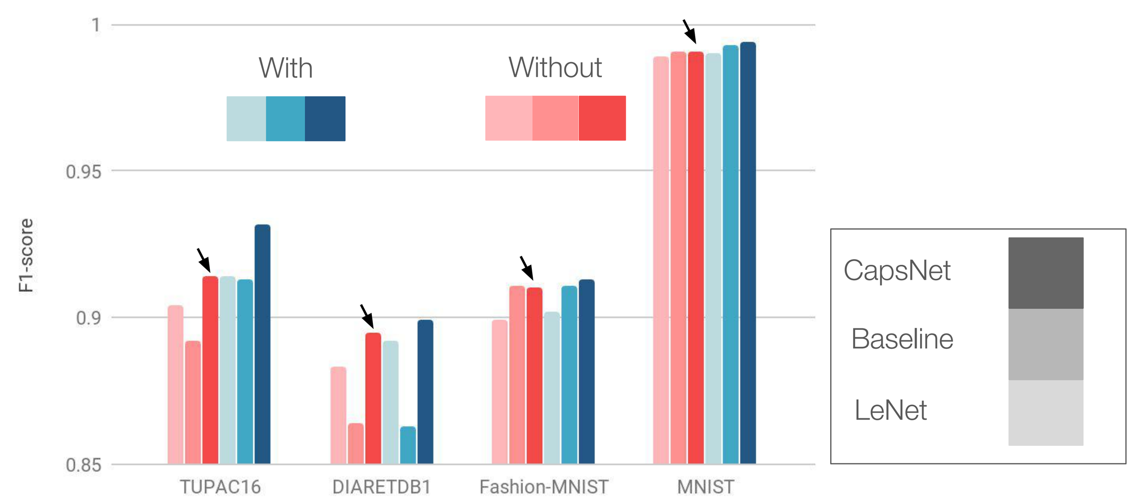


Figure 3. Mean F_1-score with and without data augmentation.

- CapsNet needs **less images** for a better performance.
- Improvement is **limited** in more complex dataset.
- All our experiments validated the significance test with a p-value < 0.05 (except for TUPAC16).

- CapsNet is **more robust** to imbalance in the class distribution.
- At least one of the imbalance cases verified the significance test for all datasets.

- CapsNet learns **a stronger representation** with less variability of the data.
- All results were found significant.

CONCLUSIONS

+ **Equivariance** requires to see fewer viewpoints of the instance of interest.
+ **Fewer parameters** for a similar/better performance.
+ CapsNet improves CADx **performance** under medical data challenges.

- Routing-by-agreement is **slower** than backpropagation (≈ convergence time).
- Improvement is **limited** in more complex datasets (TUPAC16).
- **Reconstructions** are blurry for medical datasets with complex backgrounds.

OUTLOOK

➡ Fully convolutional **decoder** to handle complex backgrounds.

➡ Explore CapsNets in a **semi-supervised** or **unsupervised** framework.

➡ Investigate the latent space to improve **explainability** and **interpretability**.

➡ Look into more suitable **medical datasets**, in which neighborhood structure plays a role for diagnosis.

ACKNOLEDGEMENTS

Obra Social "la Caixa"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713673. Amelia Jiménez-Sánchez has received financial support through the ''la Caixa'' INPhINIT Fellowship Grant for Doctoral studies at Spanish Research Centres of Excellence, ''la Caixa'' Banking Foundation, Barcelona, Spain.

REFERENCES

[1] Sabour, S., Frosst, N., & Hinton, G. E: Dynamic Routing Between Capsules. In Advances in Neural Information Processing Systems 30, pp. 3856–3866. Curran Associates, Inc. (2017).
[2] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (Nov 1998).

Universitat Pompeu Fabra Barcelona
BCN MEDTECH
CENTRALE NANTES