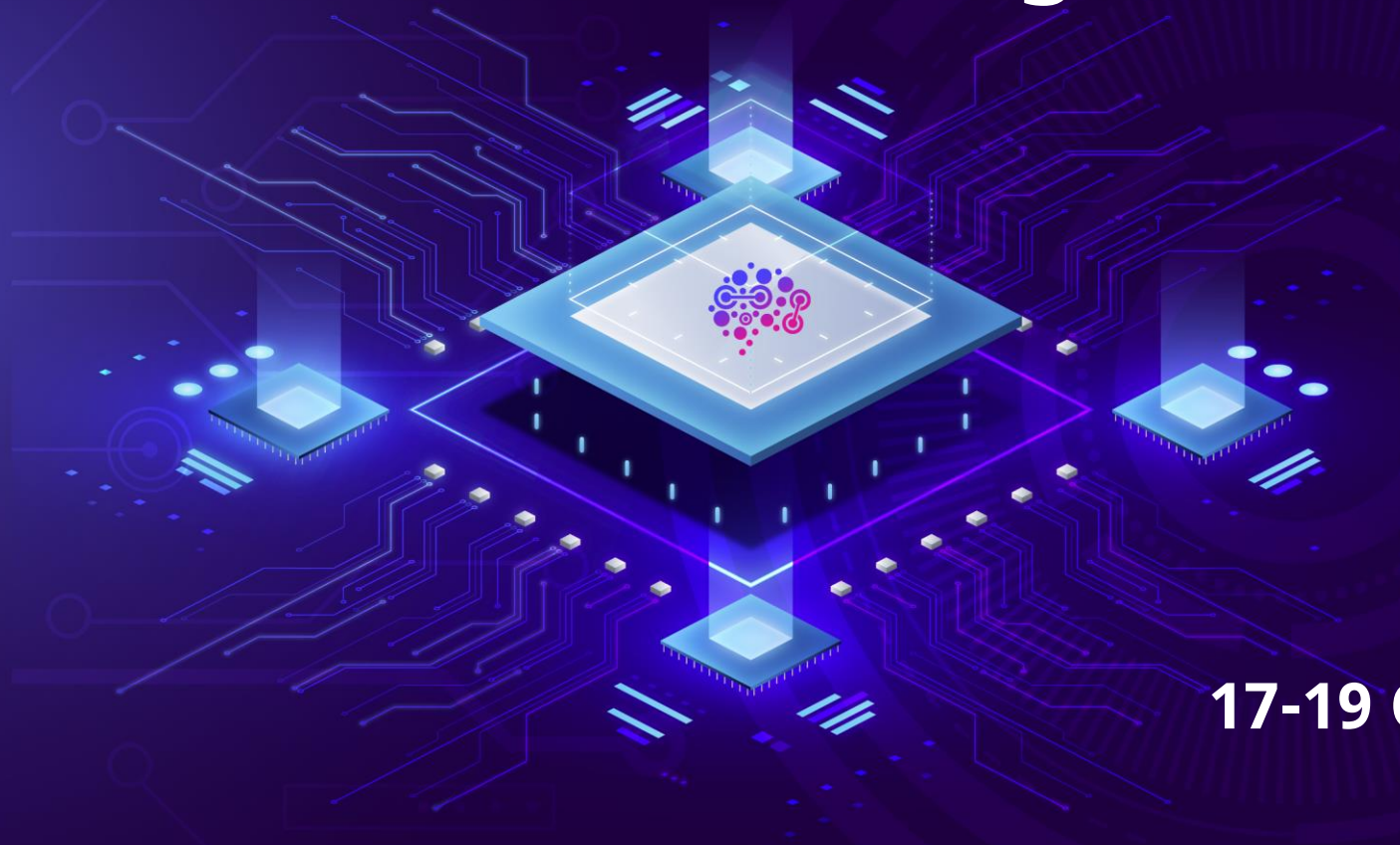
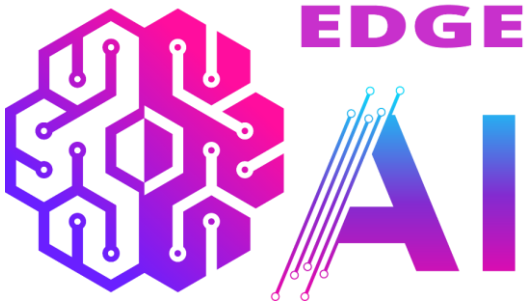
European Conference on EDGE AI Technologies and Applications - EEAI



Advancing emerging edge AI technologies and driving next generation intelligent applications.

17-19 October 2023, Athens, Greece

European Conference on EDGE AI Technologies and Applications - EEAI



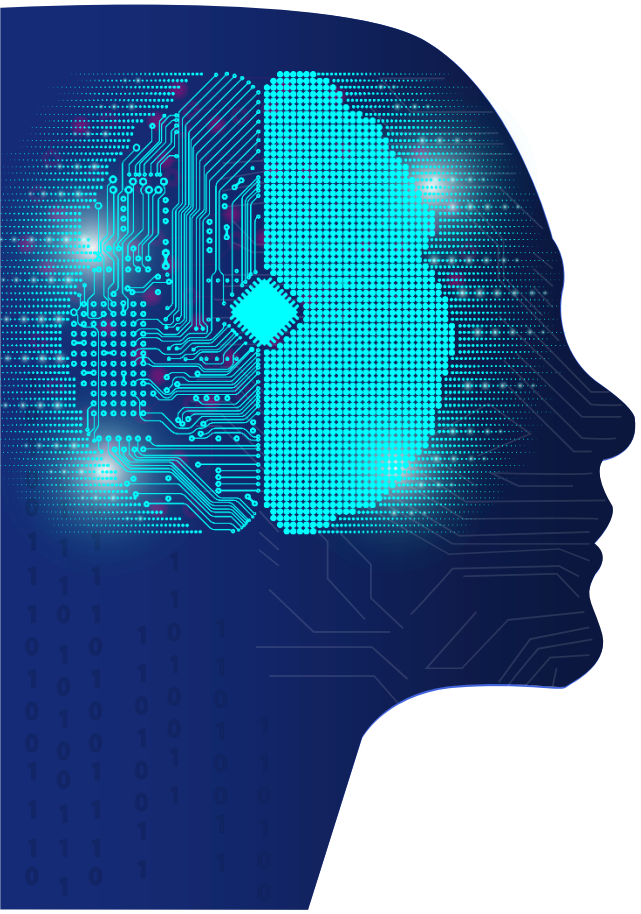
Designing Lightweight CNN for Images

HOLLARD Lilian, URCA [LICIIS], France

17-19 October 2023 Athens, Greece

Copyright © 2023

Presentation Outline



- Introduction
- State-of-the-art
- Research Goal
- Methodology
- Design and Implementation
- Experimental Results
- Future Research
- Conclusion

Copyright © 2023

Introduction

- Require high amount of data processing, memory and computing power
- Offloading computer-vision tasks to “upper” layers is difficult
  - Computer-vision data volume is far superior to more traditional IoT sensor data
  - 640p resolution video stream at 25 frames per second: 92,16 Mo / s
    - [(Width * Height * Channels) * Bytes] * fps

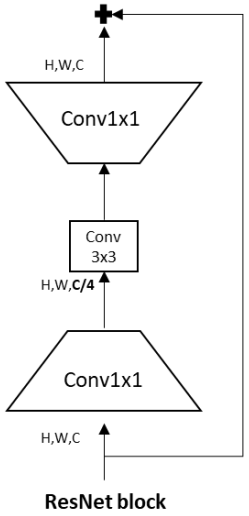
Copyright © 2023

State-of-the-art

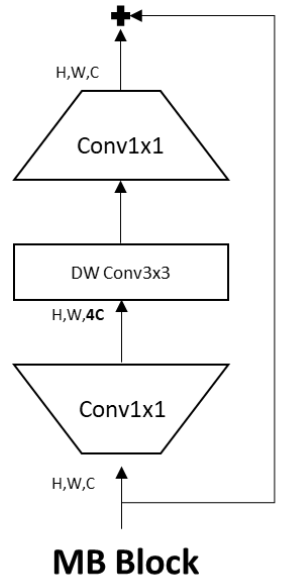
Classic advances in computer-vision methodologies for mobile devices:

- Quantification
- Pruning

- Why optimize neural network architectures?
  - Look for better training results while reducing the computational needs
  - MobileNetv2: -83.425% of FLOPS for +3.194% of accuracy improvements



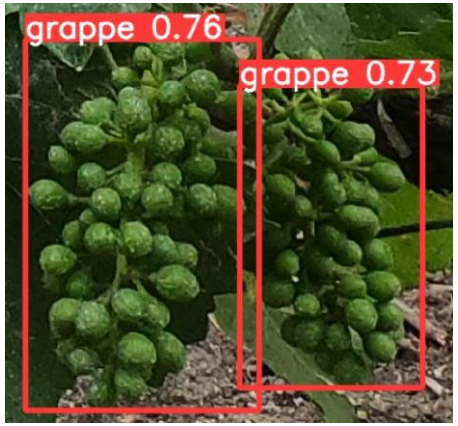
- Inverse the bottleneck : Going from low to high dimension
- Use larger filter size (which result in a larger receptive field) on more channels
  - Reduce the computational need by applying the filter depthwise

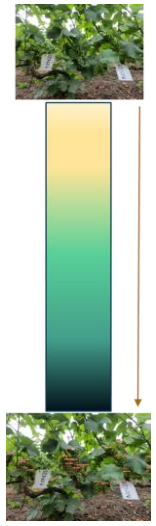


Copyright © 2023

Research Goal

- Optimizing deep neural networks for computer vision
- Reducing the computational resources of well-established deep learning models while preserving the accuracy
- Projects involving optimized models for computer vision can serve as concrete examples of use cases for our research:
  - Grapevine yield prediction
  - Grapevine disease identification

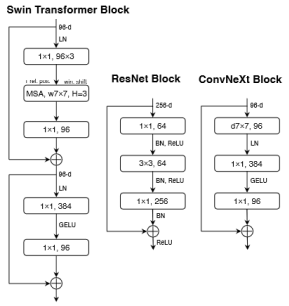


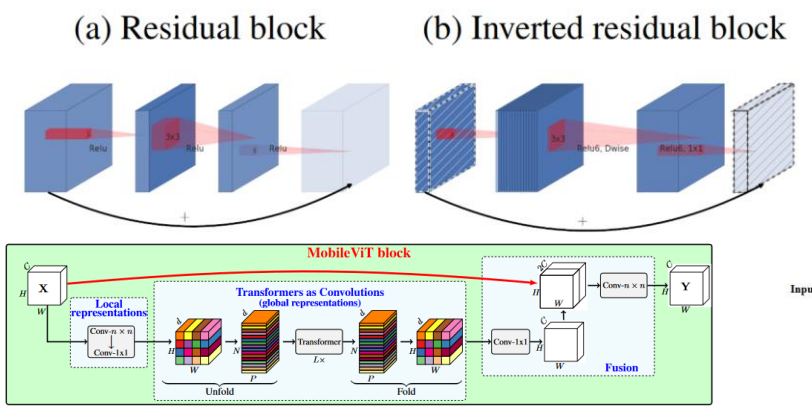


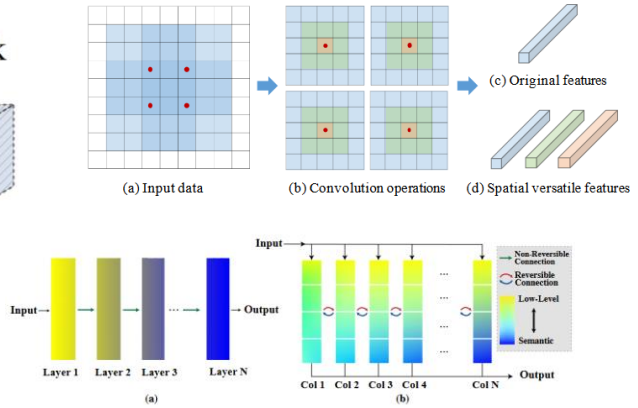
Copyright © 2023

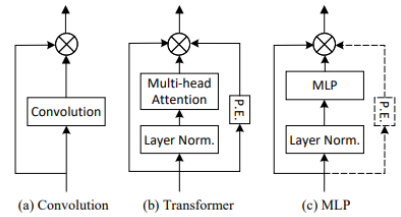
Methodology

- Looking at the evolution of deep neural network design
  - Room for optimization
  - “Modern” architectures find better accuracy
    - Better training recipes
- We focus on a vast spectrum of model architecture
  - Comparing their computational efficiency and accuracy
- Investigate similarity within the literature
  - Unwilled optimization: interpret similarity and the accuracy variation









And much more…

Copyright © 2023

Design and Implementation

- Architecture research must take care of your own needs
  - Architecture optimization = Great adaptability
- Let's illustrate with grapevine detection!
  - Grapevine Dataset repartition:

```
small: 800 (47%) - medium: 893 (53%) - large: 1 (<1%)  - Total = 1694
```

MS Coco Dataset case :

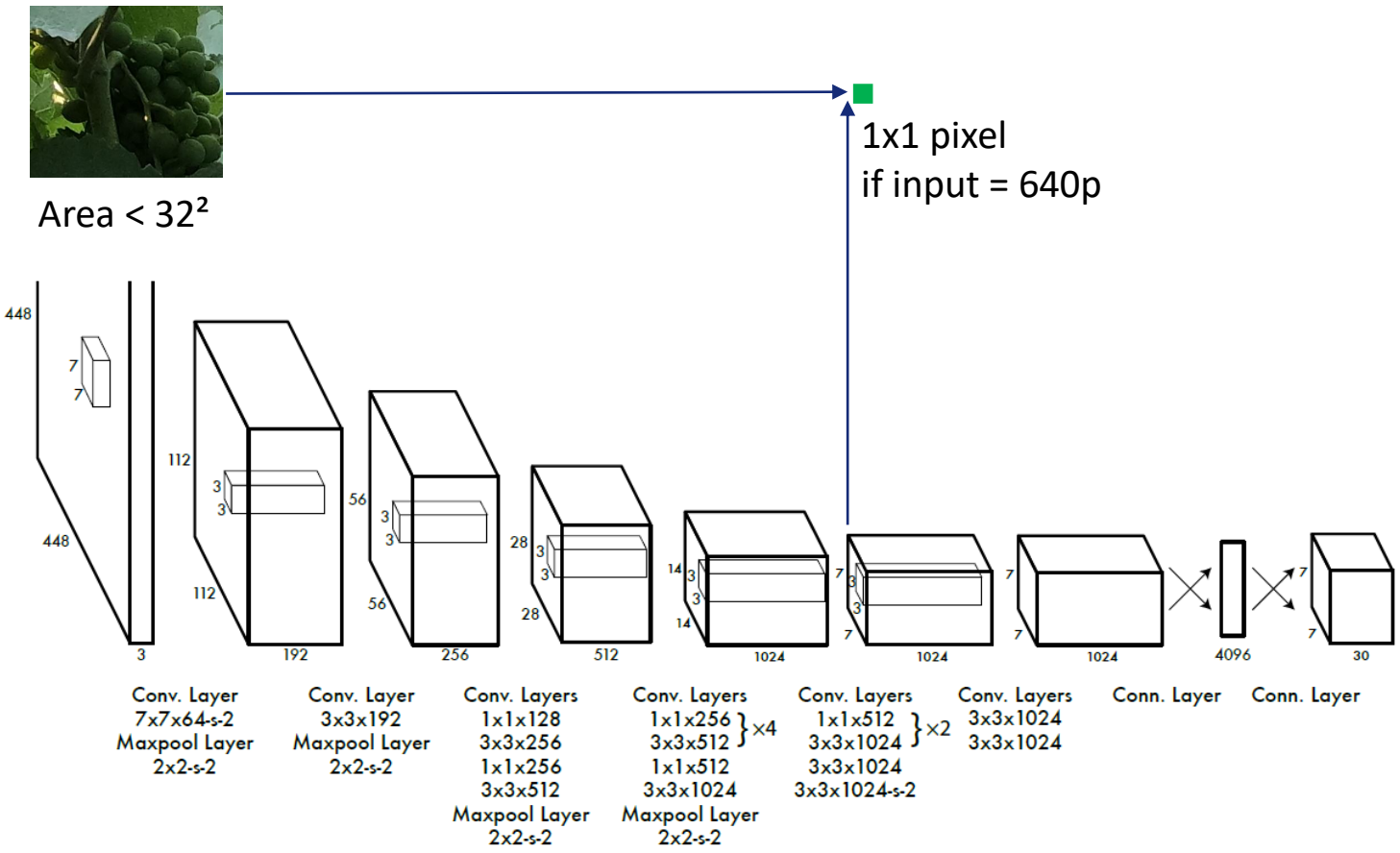
- 41% Small – $< 32^2$
- 34% Medium $< 96^2$
- 24% Large

Copyright © 2023

Design and Implementation

- Yolov1, Tiny Yolo, Tinyssimi YOLO, …
  - Why these architecture are not viable

Area < 32^2



1by1 pixel may contain erroneous information !

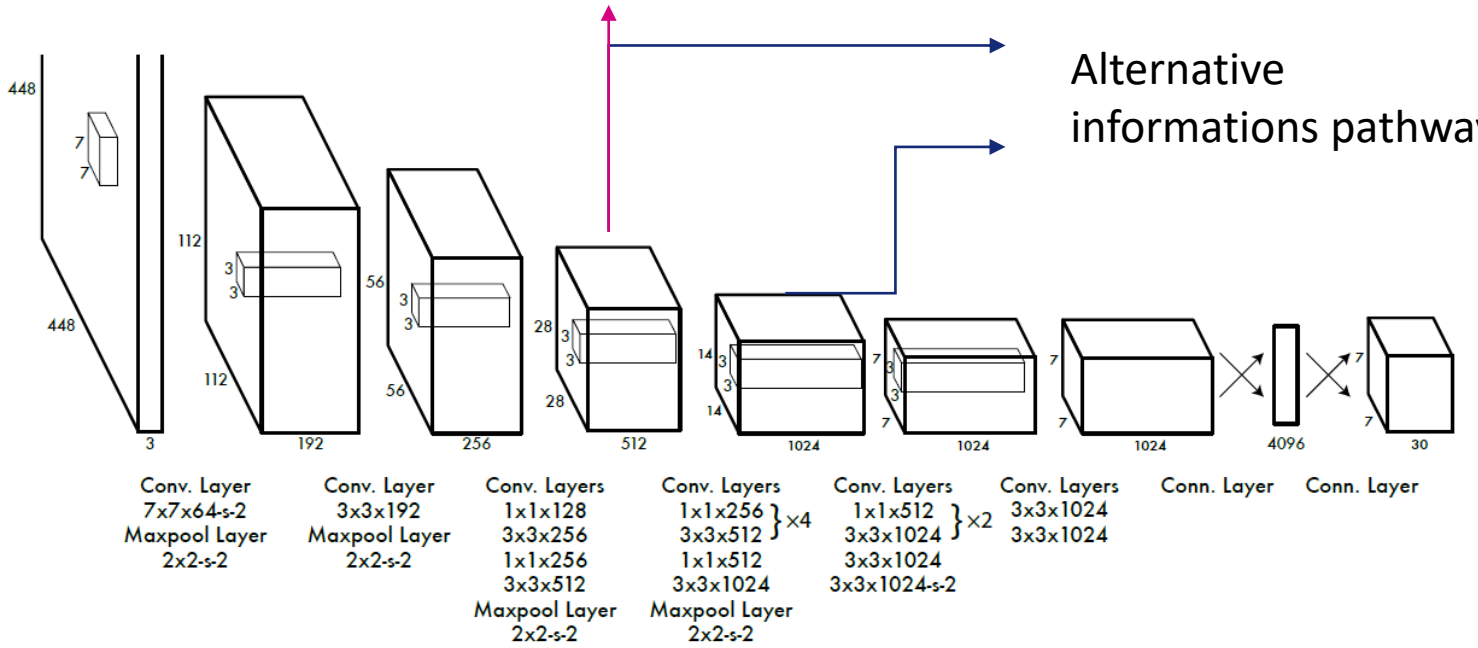
- For training…
- .. and inference purpose

Copyright © 2023

Design and Implementation

- Yolov1 – Why this architecture is not viable anymore
  - Grapevine case study

Area < 32^2



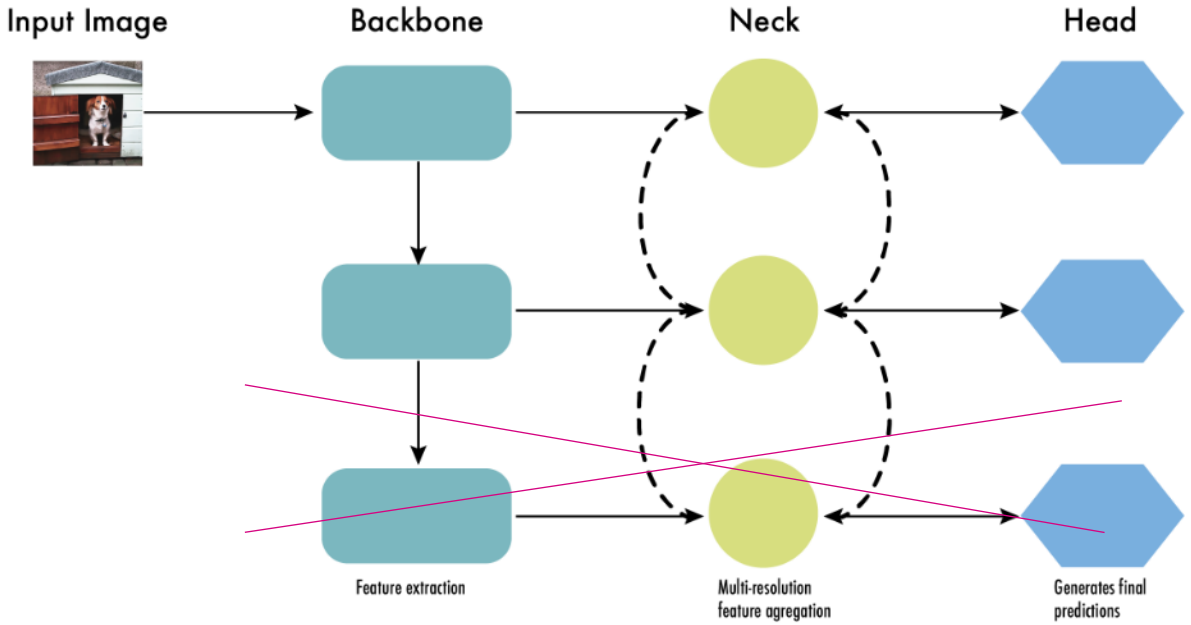
Some architectures with multiple head outputs :

- FPN
- PANet
- **CSPPanNet (heavily used in YOLOs)**
- EfficientDet – BiFPN
- NAS-FPN
- FC-FPN
- …

Copyright © 2023

Experimental Results

- Yolov5 is better than Yolov4, v3 **for grapevine detection** at nano level
  - Best scale/accuracy ratio
  - And faster than EfficientDet with BiFPN detection neck



Latest Yolov5 version use :
- 4 detection head (Xlarge object)
- Equivalent accuracy with 2 detection head

Yolov7 Tiny - 13,7G - 6,2M
Yolov5 Nano - 4,5G - 1,9M – 74,4%

Our model :
Yolov5 modified version (2 neck), fewer channels in narrow layers - 1,8M – 76,6%

Copyright © 2023

Future Research

- Several architectures might be more interesting for YOLO models
  - Feature extraction :
    - **MobileNetv3**
    - HRNets
    - RepCoI
    - TinyViT
    - EfficientViT
    - FastViT
    - **MobileNeXt**
    - **Res2Net**
    - **MobileOne**

Experimental results have shown that our data respond better with a larger number of channels.

MobileNetv3 network with Squeeze Excitation block

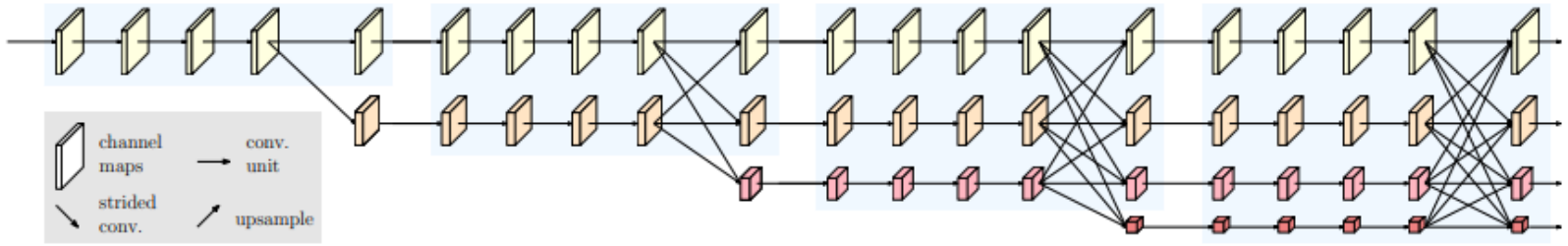
- Great flops/channels ratio per block

Copyright © 2023

Future Research

- Several architectures might be more interesting for object detector
  - Optimize the way the neck and the head is working as an object detector
    - Parallelization nightmare
    - Gathering several semantical information induces nonhomogeneous waiting time, different sizes of activation maps to keep in memory, nonhomogeneous repartition of computational time

  - HRNet backbone example:
    - Splitting and gathering the information between several semantic levels is done on the same waiting process



Wang, Jingdong, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, et al. « Deep High-Resolution Representation Learning for Visual Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, nº 10 (octobre 2021): 3349-64. https://doi.org/10.1109/TPAMI.2020.2983686.

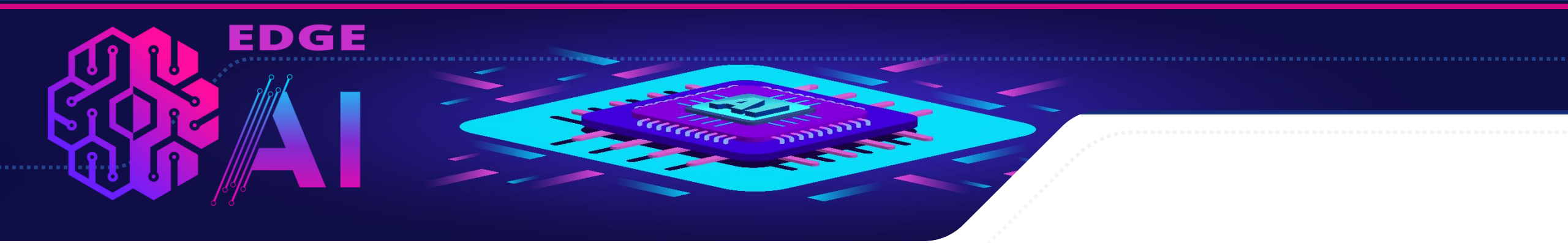
Copyright © 2023

Conclusion

- Essential contribution to the Computer vision community as a whole
  - Architecture optimization induces a vast computational scaling use case
  - Architecture optimization might improve or stabilize a given accuracy for less computational requirements.

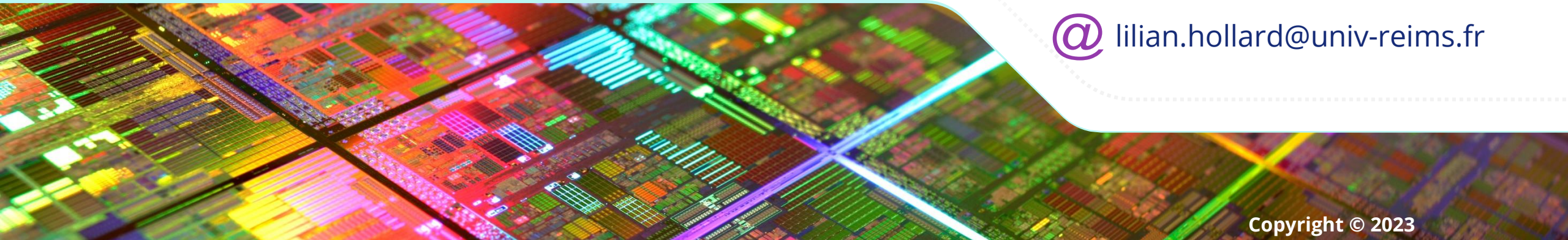
- Optimizing deep neural network architecture for computer vision is my Ph.D. thesis subject!
  - Ask me anything
  - Feel free to give me any feedback

Copyright © 2023



Thank You

For your attention



@ lilian.hollard@univ-reims.fr

Copyright © 2023