Towards more efficient data preprocessing in deep networks

Fernando Peña Bes[1], Ana C. Murillo, Darío Suárez Gracia
DIIS, I3A, Universidad de Zaragoza, Spain

ACACES 2023

Problem

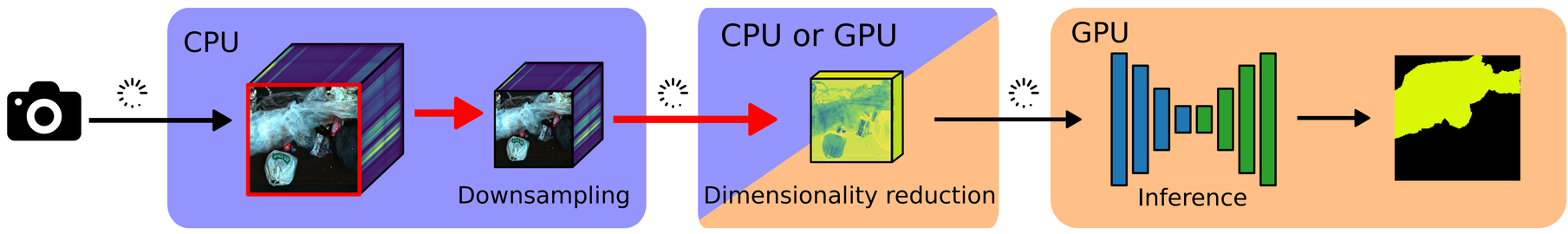
- Using **highly dimensional** data with deep learning models is challenging due to its complexity and size of the datasets
- We analyze the bottlenecks associated when processing **hyperspectral images** in a segmentation network
- As a practical example, we address an **object detection** problem within the **recycling sector**

Setup

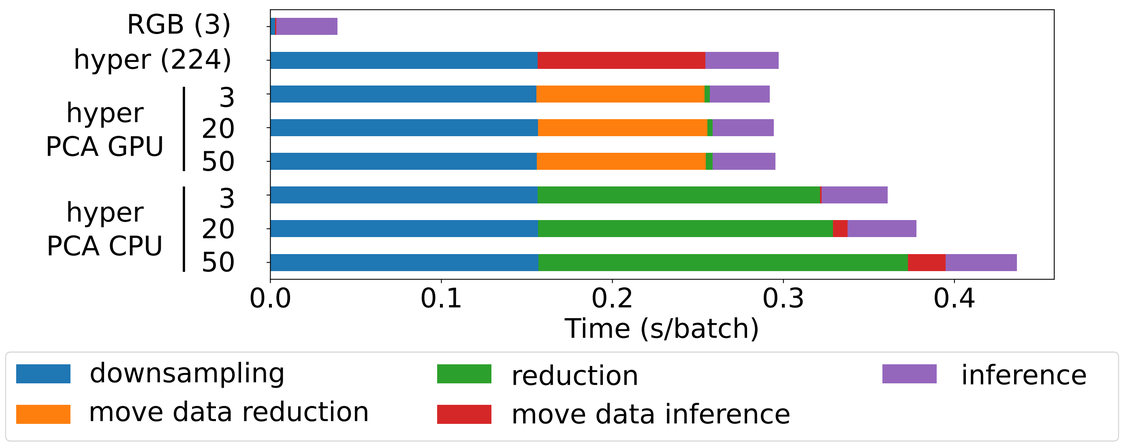
The dataset is composed of 222 scenes from a recycling line, with 377 segmentation masks denoting big plastic objects that can cause obstructions in the line. Each scene was captured with both an **RGB** and a **hyperspectral** camera.



We use MiniNet-v2 [1] as baseline model combined with PCA for dimensionality reduction varying the number of components.



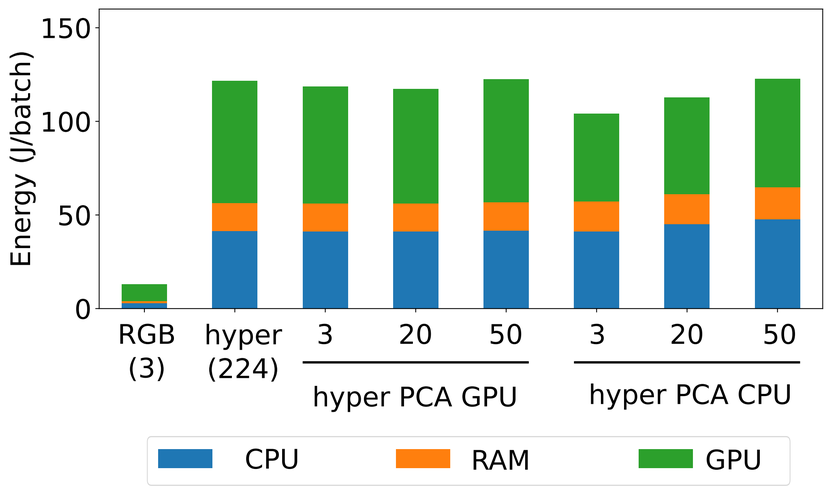
Performance results

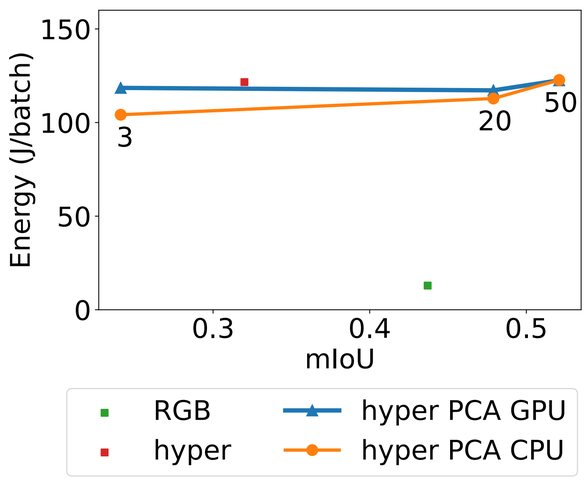


Processing time

When using hyperspectral images most of the time is spent in **preprocessing steps** (downsampling and reduction) rather than inference because of the higher dimensionality (224 spectral channels vs. 3 RGB channels).

Performing PCA on the CPU allows to **reduce transfer times**. However, the extra time spent in the reduction increases the global time.





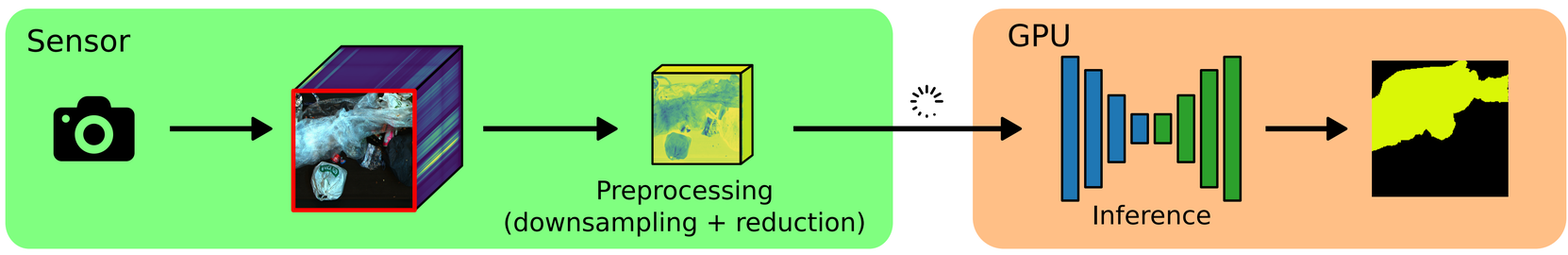
Energy consumption and accuracy

Dimensionality reduction facilitates training as it **compresses redudant information** present in hyperspectral images.

Also, although performing PCA on the CPU is slower, there is a **slight reduction in the total energy consumption** because the CPU consumes less energy than the GPU.

Conclusions

- The two **principal bottlenecks** in the pipeline are the **preprocessing** steps and **data transfers** between devices
- **Efficient preprocessing** techniques to represent the data in a reduced form are crucial to minimize the data transferred
- They can be incorporated in **hardware accelerators close the sensors** [2], delivering data that can be efficiently handled in the rest of the pipeline



Grupo de Investigación en Arquitectura de Computadores (gaZ)
Universidad Zaragoza
1542

References

[1] I. Alonso et al. MiniNet: An efficient semantic segmentation ConvNet for real-time robotic applications, *IEEE Transactions on Robotics*. 2020

[2] W. Dally. On the model of computation: Point, *Commun. ACM, 65(9)*. 2022

Link to the poster abstract

[1]**Contact:** ferp@unizar.es