From Few-Shot to Zero-Shot Pallet Load Recognition: A Deployed Embedding-Based Vision System for Industrial Logistics Supplementary Material

Juan Jesús Losada del Olmo[1,2], Emilio Pardo Ballesteros[1], Pedro E. López-de-Teruel[1,2], Alberto Ruiz[1,2]
[1]University of Murcia, Spain [2]Bleecker Technologies, Spain
{juanjesus.losada, emilio.pardob, pedroe, aruiz}@um.es

In this supplementary material, we provide additional details regarding our experimental setup and hyperparameter tuning to ensure the reproducibility of our work. We begin in Section A by outlining the hardware specifications used for all experiments. In Section B, we detail the set of image resolutions studied. Subsequently, we present two parameter tuning studies that justify our methodological choices: Section C analyzes the impact of PCA dimensionality reduction on the performance of our few-shot approach, while Section D evaluates the impact of the number of neighbors used, k, on the accuracy of the k-NN classifier.

A. Hardware specifications

The hardware configuration for all experiments consisted of an `Intel Core i9-14900KF 3.2/6 GHz` CPU and the `NVIDIA GeForce RTX 4070 Ti 12 GB` GPU.

B. Experimental image resolutions

The original image resolution of the images in the public dataset used for our experiments is 2688×1520 pixels. However, limited by our hardware's capacity for the vision foundational models under consideration, we downscaled the images to 1064×602 pixels, the largest resolution divisible by 14×14. Furthermore, lower-resolution images were also studied as a means to both accelerate the inference process and determine how models would perform under such conditions. All resolutions considered are listed in Table 1.

C. Impact of PCA dimensionality reduction

Figure 1 qualitatively illustrates how the number of principal components used in PCA reduction impacts the performance of our few-shot approach. Specifically, the figure presents two scenarios: a one-shot setting (top) and a five-shots setting (bottom). Both are evaluated in a multiclass context, though the observed behavior is analogous in a binary context, which simply yields slightly higher accuracy.

Table 1. Image resolutions and the count of 14×14 tiles per case.

Resolution (H × W)	#Tiles (H/14 × W/14)
1064×602	$76 \times 43 = 3268$
854×476	$61 \times 34 = 2074$
686×378	$49 \times 27 = 1323$
546×308	$39 \times 22 = 858$
434×238	$31 \times 17 = 527$
350×196	$25 \times 14 = 350$

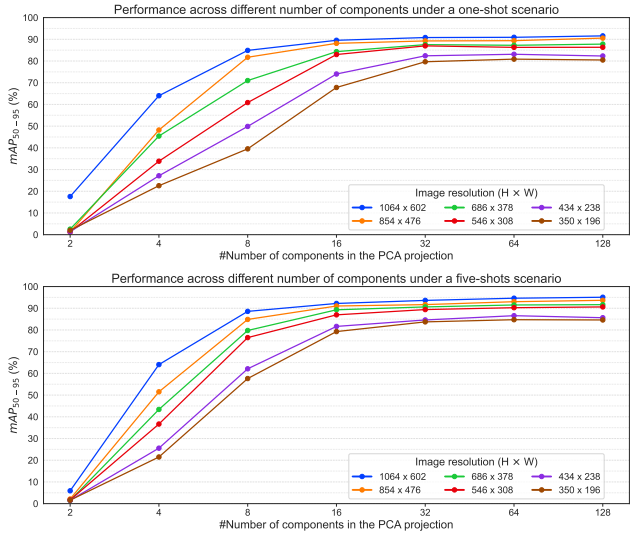

Figure 1. **Effect of PCA dimensionality on multiclass detection.**

Across both scenarios, the graphs show that for all image resolutions, 32 principal components generally represent a saturation point; further increases in dimensionality toward the full 1024 raw CAPI features offer no subsequent performance gain. Under the one-shot scenario, a component count of 16 appears sufficient for the top two resolutions, while 32 components are needed for the remaining reso-

lutions to achieve stable performance. Conversely, in the five-shots scenario, which represents the largest support set in our study, the four highest resolutions appear to stabilize with just 16 components, whereas the two lowest resolutions still demand a higher number of components (~ 32).

Therefore, after analyzing these two cases, we decide to set the number of principal components for our few-shot studies to $n_{\text{pca}} = 30$, a value near the saturation point of 32.

Additionally, a key advantage of this dimensionality reduction is the substantial decrease in the memory footprint of the system's memory bank. Specifically, reducing the feature space from 1024 to just 30 components results in an approximate 34-fold reduction in storage size (since $1024/30 \approx 34$). This substantial memory saving is particularly beneficial for the final deployment, as it not only minimizes data storage requirements but also streamlines the maintenance of the dataset. Furthermore, this compact representation enables more efficient implementations of the k-NN classifier, serving to boost overall system scalability.

D. Impact of k on k-NN performance

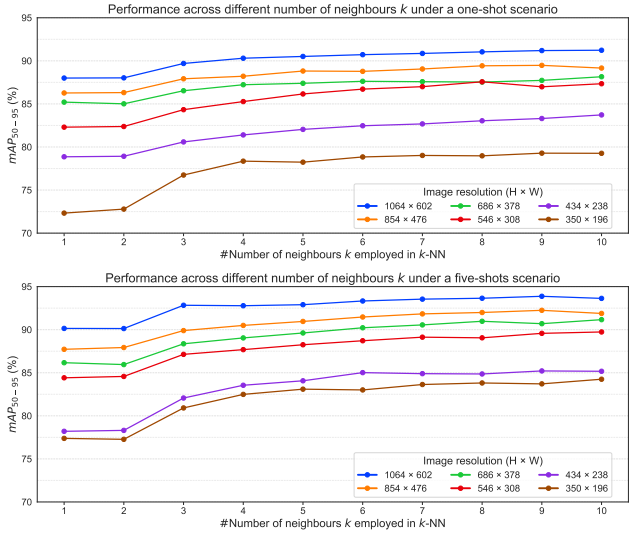

Figure 2. **Impact of k value in k-NN on multiclass detection.**

Following the analysis of PCA dimensionality, Figure 2 also depicts the influence of the k-NN classifier's neighbor count (k) on our few-shot method's accuracy. The plot contrasts two settings, a one-shot (top) and a five-shots (bottom) configuration, both in a multiclass regime. As before, these trends are representative of the behavior in a binary setting.

For both the one-shot and five-shots scenarios, the accuracy curve shows a steep ascent up to $k = 4$, suggesting significant benefits from adding neighbors in this range. However, the plot reveals diminishing returns, with performance gains becoming marginal beyond $k \in [5, 7]$. A comparable trend is noted in the two scenarios we examined, an observation that seems independent of the support set dimension.

Based on this analysis, we selected $k = 7$ for our few-shot experiments. By selecting this value, we aim to capture the majority of the accuracy improvements without the computational cost of an unnecessarily large neighbor set.