On the Minimal Recognizable Image Patch

Mark Fonaryov **Michael Lindenbaum**

Technion - Israel Institute of Technology

markf@campus.technion.ac.il **mic@cs.technion.ac.il**

Abstract

In contrast to human vision, common recognition algorithms often fail on partially occluded images.
We propose characterizing, empirically, the algorithmic limits by finding a minimal recognizable patch (MRP) that is by itself sufficient to recognize the image. A specialized deep network allows us to find the most informative patches of a given size, and serves as an experimental tool. A human vision study recently characterized related (but different) minimally recognizable configurations (MIRCs) [*], for which we specify computational analogues (denoted cMIRCs). The drop in human decision accuracy associated with size reduction of these MIRCs is substantial and sharp.
Interestingly, such sharp reductions were also found for the computational versions we specified.

[*] – Shimon Ullman, Liav Assif, Ethan Fetaya, and Daniel Harari. **Atoms of recognition in human and computer vision.** Proceedings of the National Academy of Sciences, 113(10):2744–2749, 2016.

Motivation

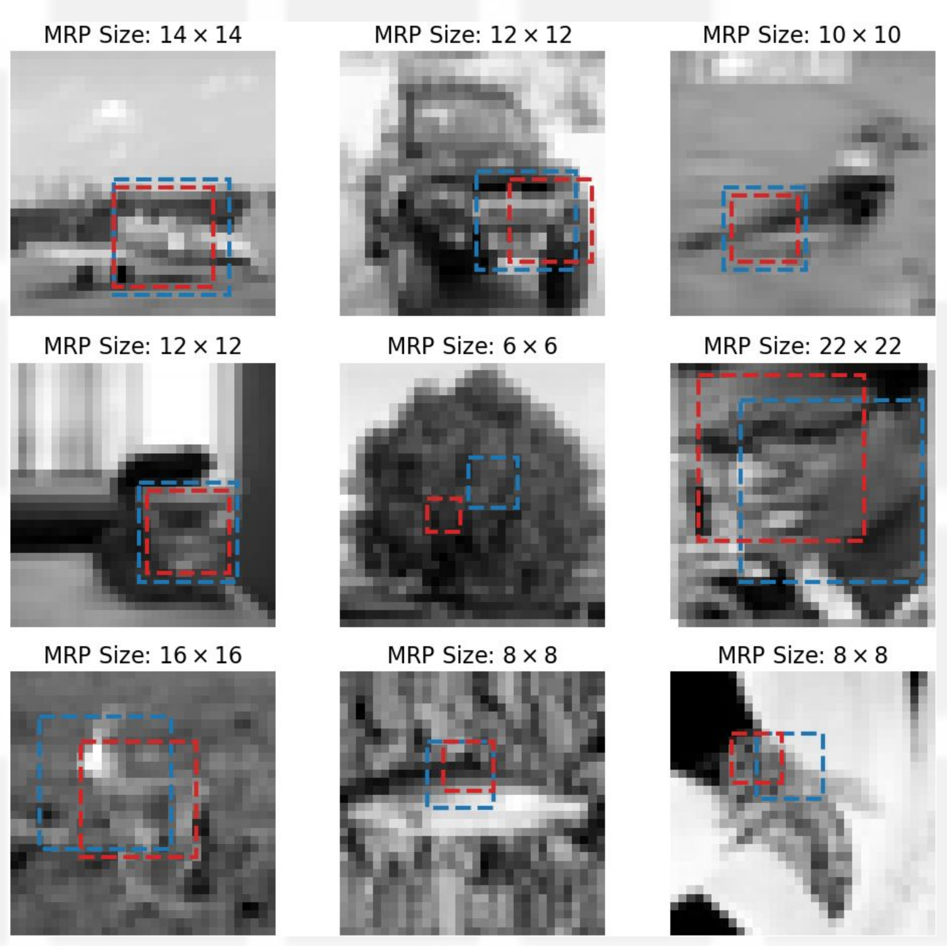
Recognition from parts – both practical and interesting:

- Objects behind occlusions – recognition from incomplete visual information.
- Objects behind occlusions – requires features based on object parts.

Partially occluded objects are very challenging for recognition algorithms.

- Algorithmic recognition – quickly deteriorates with partially occluded or cropped objects.
- Human recognition – works well with very limited visual information.

Natural question: what is the minimal image-patch that suffice for object recognition?



Patch Recognizability

Locally Recognizable Patch - a patch that is correctly categorized by the classifier.
q-Locally Recognizable Patch - correct categorization with a confidence higher than q.

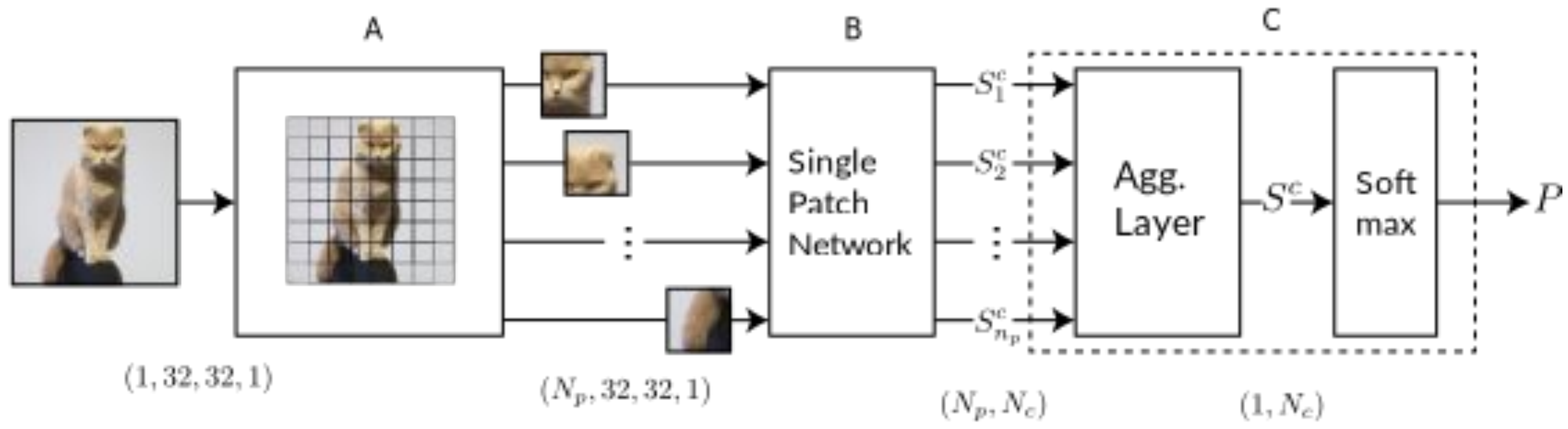
$$\hat{c} = argmax_c \; S_p^c \qquad (1)$$

Globally Recognizable Patch - the correct score associated with this patch is higher than all other scores associated with all other patches and other classes.

$$\hat{c} = argmax_c \; (max_p \; S_p^c) \qquad (2)$$

In our study, class scores are provided by a CNN - the single-patch-network (SPN).

Patch Based Classification



(A) Input image is split into N patches, each resized to 32x32.
(B) Each path passes thru the single-patch-network (SPN).
(C) Aggregation - patch-level scores -> image-level scores -> image-level probabilities.

Patch Score Aggregation

Aggregation influences the confidence associated with the different categories.
Category-independent max - maximum score is evaluated separately for each category.
Winner-directed max - scores are taken from a single patch with the overall highest score.

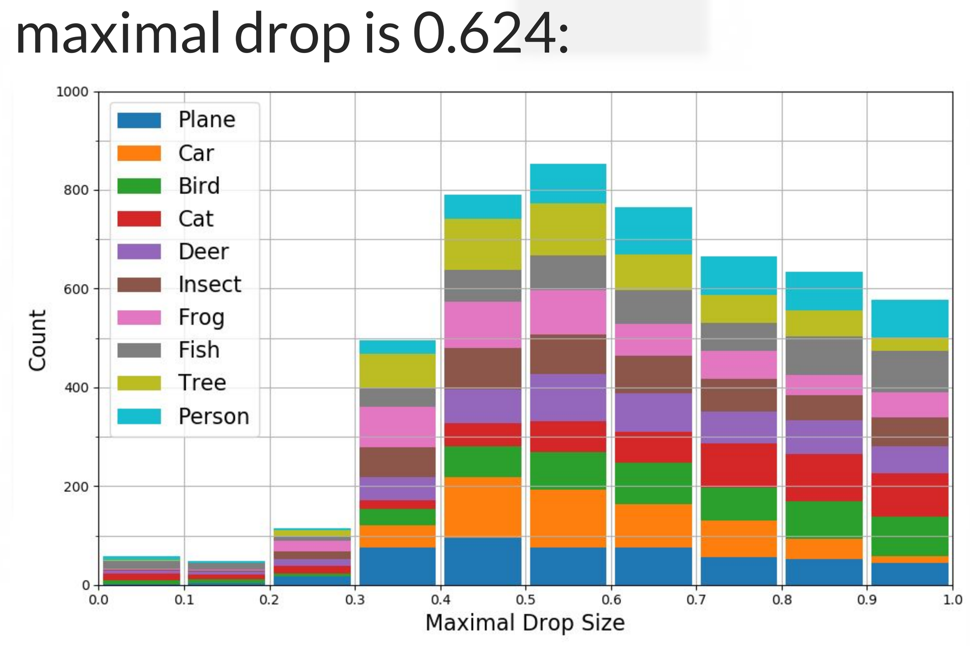
$$S^c_{max-ind} = \max_p \{S^c_p\}, \qquad (3)$$

$$S^c_{max-dir} = S^c_{p^*} \;, \text{where } p^* = argmax_p \; (max_c \; S^c_p) \qquad (4)$$

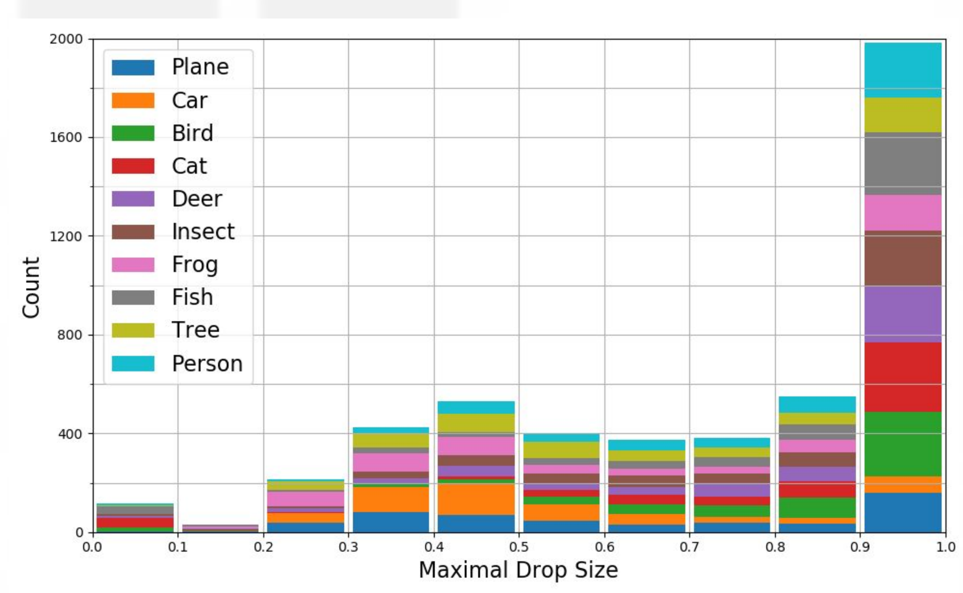
Single Image Recognizability

Histograms of maximal confidence drops:

- Most images include a sharp and significant sharp drop.
- Category-independent max - average maximal drop is 0.624:

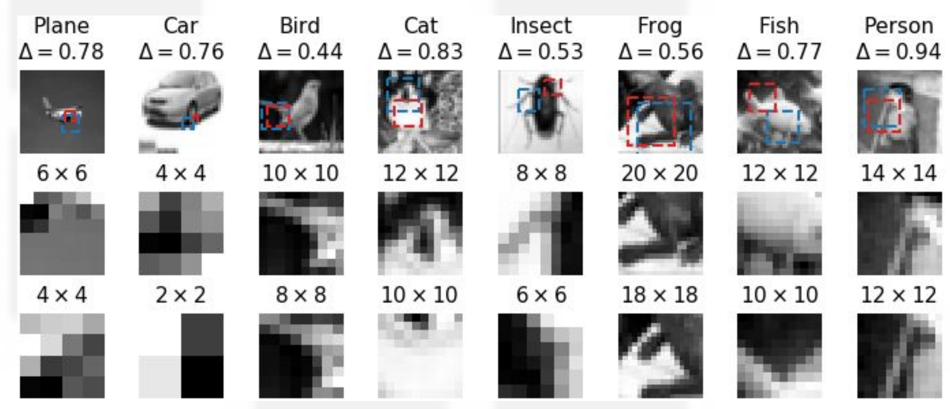


- Winner-directed max - average maximal drop is 0.72:



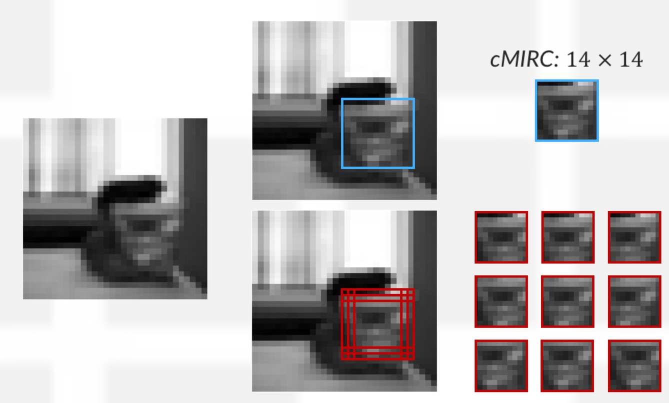
Globally Minimal Patches

We denote the minimal recognizable patch (MRP) as the minimal patch size with the correct image classification, associated with the maximal score.



Locally Minimal Patches

Computational *MIRC* (*cMIRC*) – a patch that is *q-locally recognizable*, while all its nine contained sub-patches are not.



Conclusions

Both *MRPs* and *cMIRCs* share a common property with human vision –> sharp drops.

MRPs were small, and usually unrecognizable by humans.