

A Critical Look into Quantization Table Generalization Capabilities of CNN-based Double JPEG Compression Detection

Elena Rodríguez-Lois, David Vázquez-Padín, Fernando Pérez-González, and Pedro Comesaña-Alfaro

{erodriguez|dvazquez|fperez|pcomesan}@gts.uvigo.es

EUSIPCO 2022. 29 August - 2 September 2022, Belgrade (Serbia)

Introduction

- **Problem:**
 - CNN-based detectors achieve state-of-the-art performance in double JPEG compression detection.
 - These are typically tailored to the training conditions and suffer a drop of performance when evaluating unknown sources.
- **Popular approach:**
 - Some works use a highly diverse dataset (holistic approach), to reduce dependency on the source.
- **Our goal:**
 - This work assesses the generalization capabilities of such double JPEG compression detectors w.r.t. unknown Quantization Tables.

Source Mismatch

According to [1] one can differentiate:

- **Intrinsic difficulty:** Difficulty baseline of an unknown source, as the score over a model trained on the source.
- **Inconsistency:** Difference in difficulty over a model trained on a different source.

These definitions can be expanded to accommodate sets of sources instead of individual ones.

Comparing Quantization Tables (QT)

The work in [2] introduces a QT dissimilarity semi-metric, which can be used to:

- Compare the quantization coefficients of any two QTs ($\mathbf{p}$ and $\mathbf{q}$), as

$$d^2(\mathbf{p}, \mathbf{q}) = \sum_{k,l \in \{1, \dots, 8\}} \frac{1}{(k+l)^2} \left(\frac{q_{kl} - p_{kl}}{q_{kl} + p_{kl}} \right)^2.$$

- Estimate the closest QF to a non-standard QT, as

$$\hat{\mathbf{QF}}(\mathbf{q}) = \arg \min_{\mathbf{QF} \in \{1, \dots, 100\}} (d(\mathbf{q}, \mathbf{q}_{\text{st}}(\mathbf{QF}))).$$

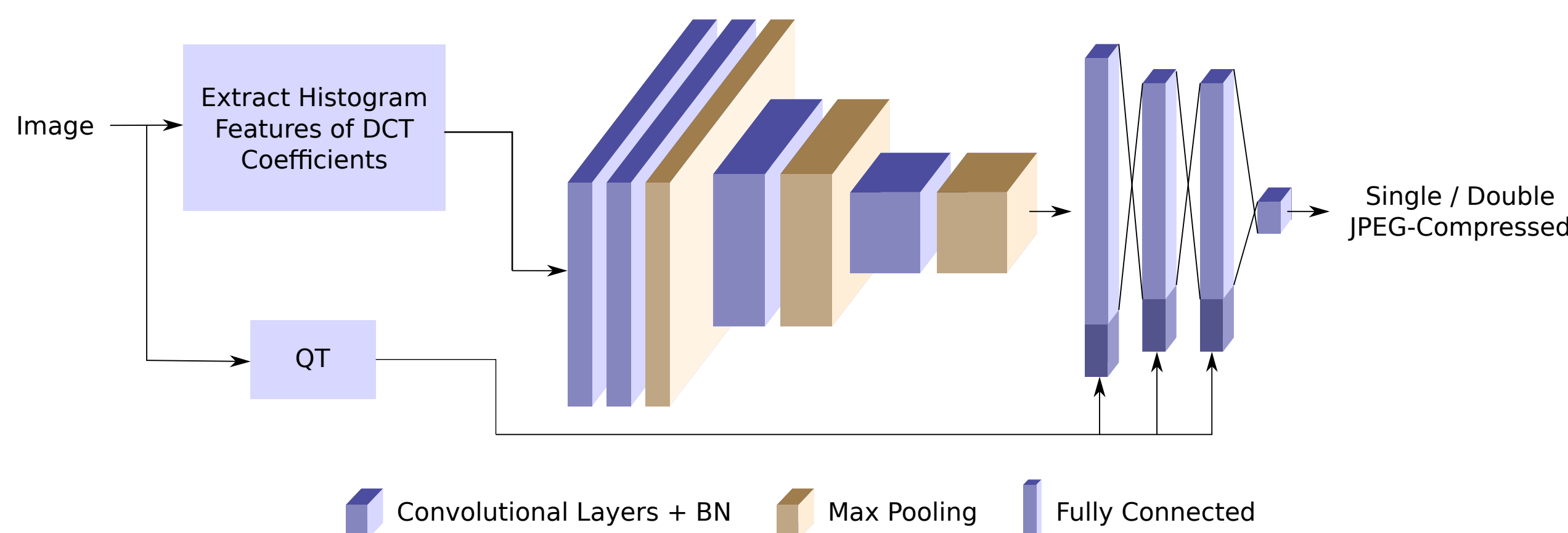
These can be used to define operating regions with similar characteristics.

Park et al.'s ConvNet

Park et al. [3] proposed a CNN-based double JPEG compression detector exploiting:

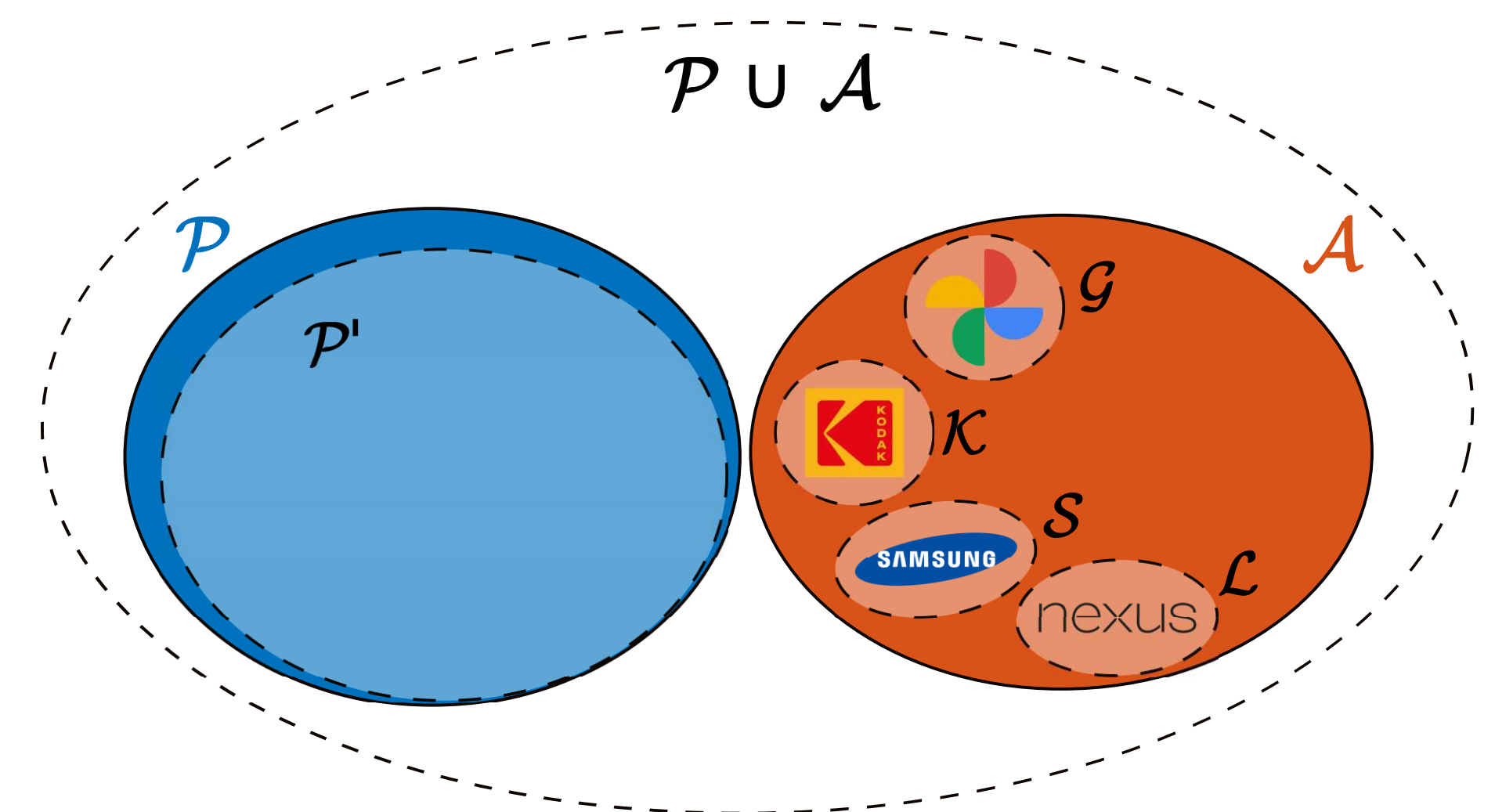
- Histogram features from DCT coefficients.
- Quantization Table side-information from the last JPEG compression.

And trained it using a highly diverse training dataset of 1120 different QTs ($\mathcal{P}$).



The performance of the detector drastically varies according to the intrinsic difficulty of the operating regions.

Quantization Tables Subsets

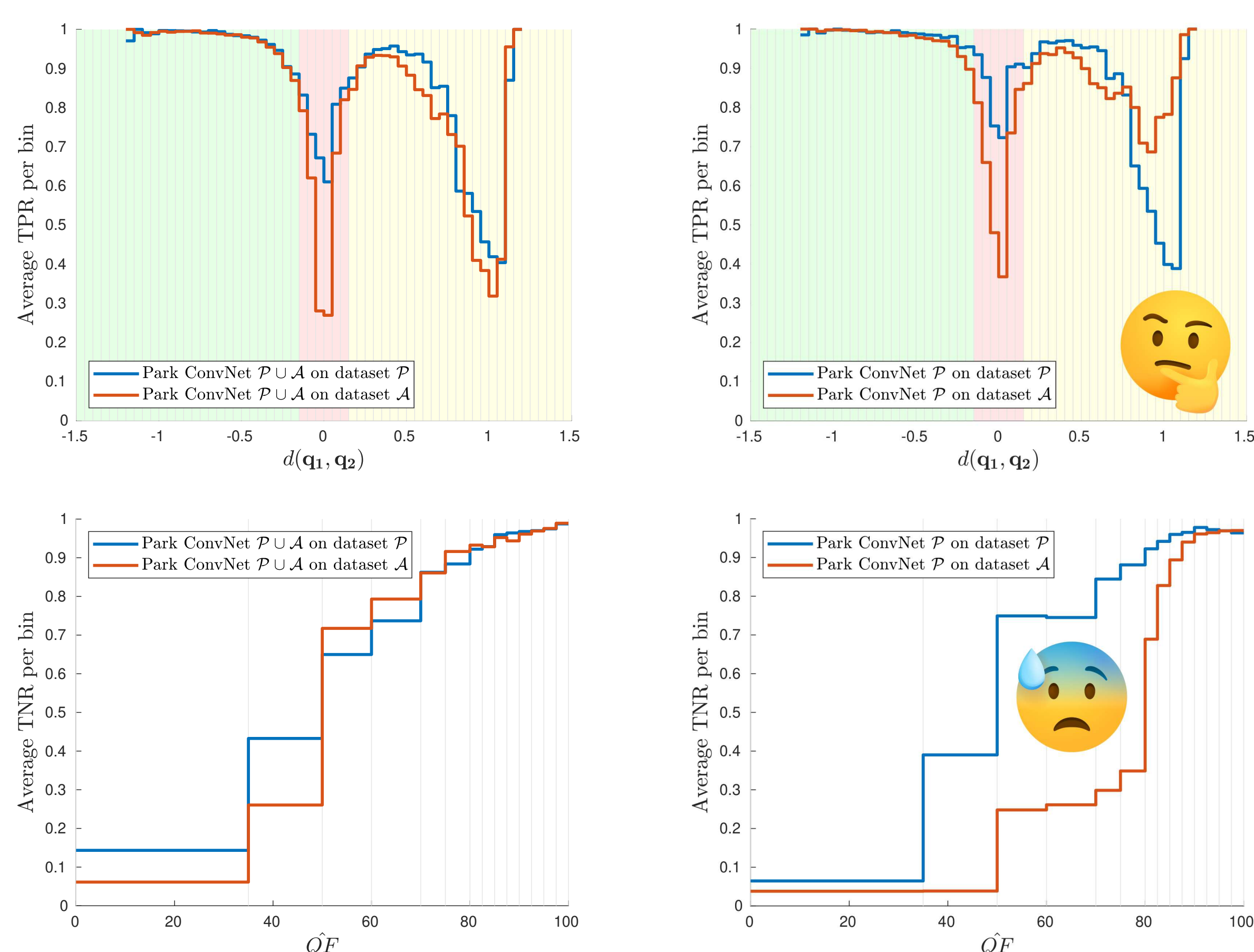


These subsets were combined and used to create:

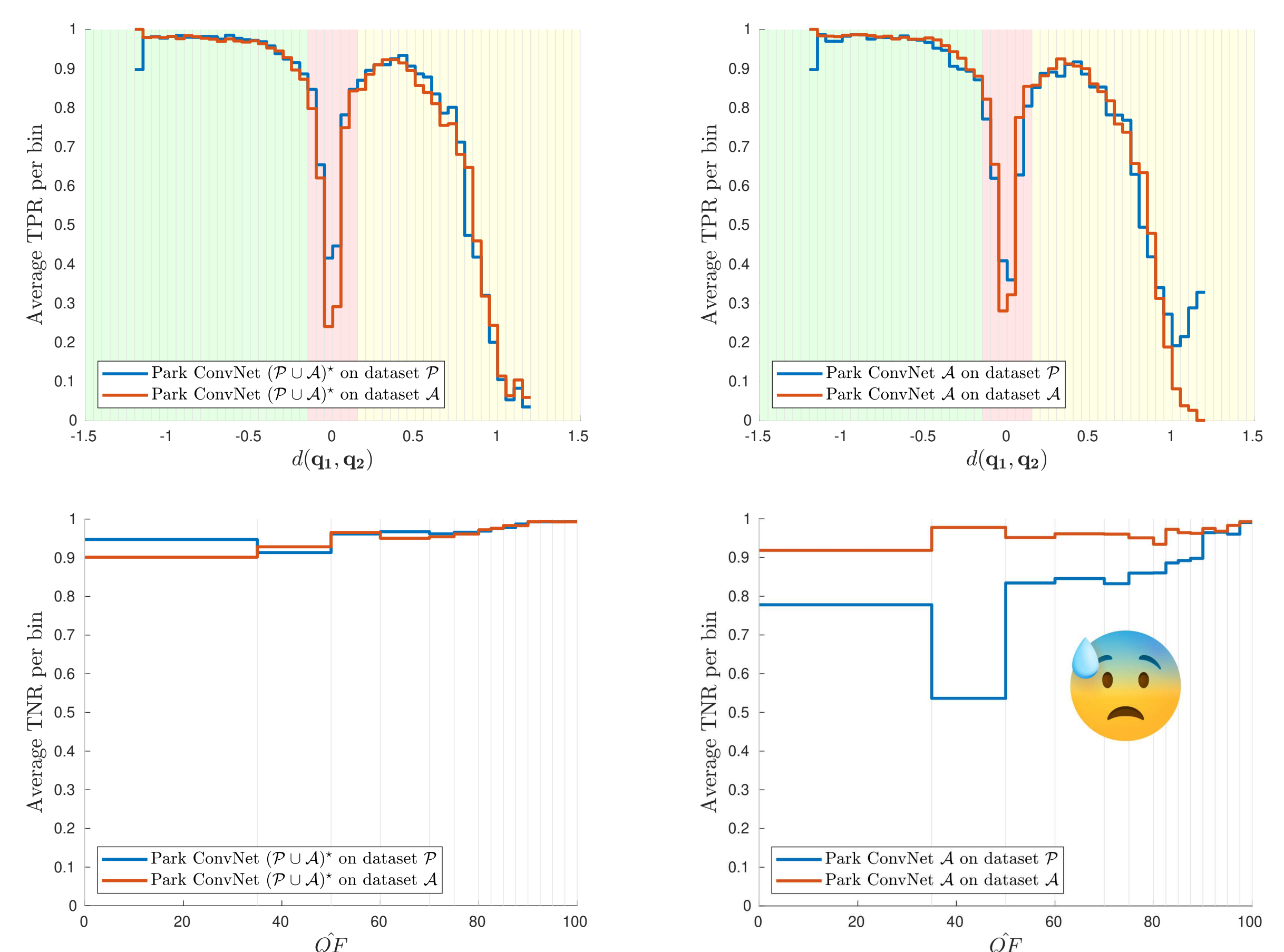
- 7 Random Datasets to train and evaluate ParkConvNets.
- 3 Pipeline Datasets to assess the inconsistency in a realistic scenario.

Experimental Results

Inconsistency Analysis: $\mathcal{A}$ w.r.t. $\mathcal{P}$



Inconsistency Analysis: $\mathcal{P}$ w.r.t. $\mathcal{A}$



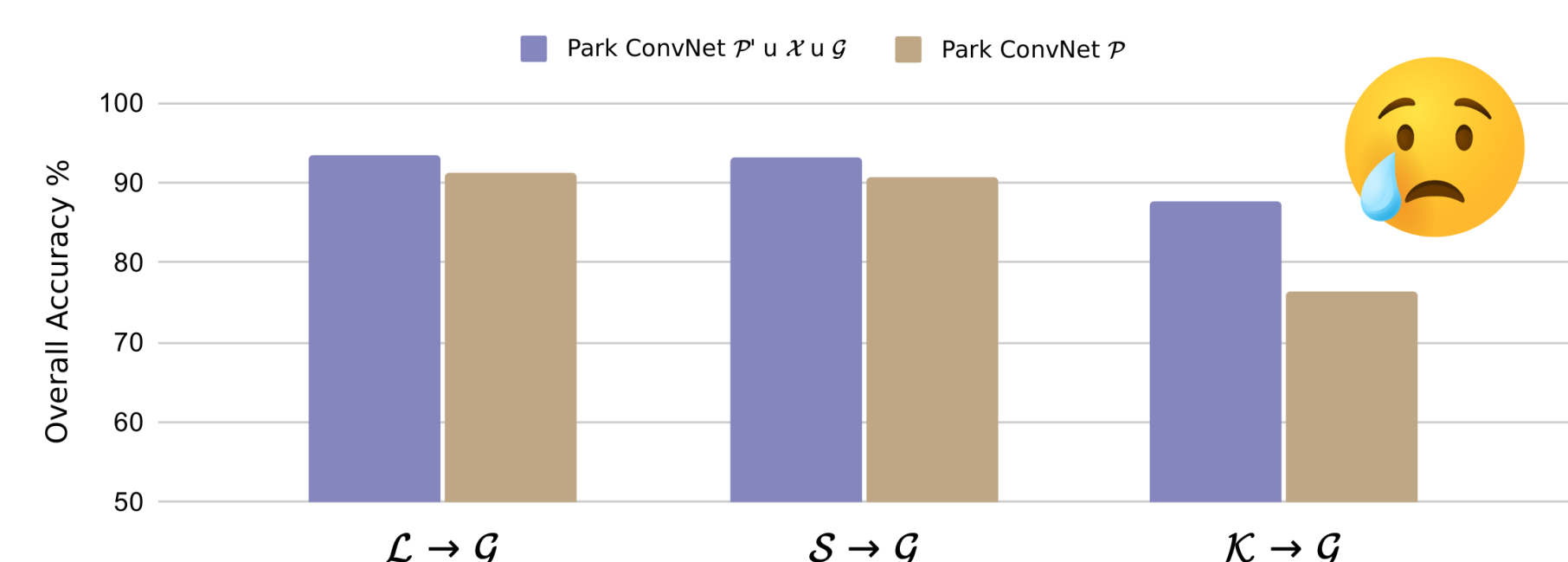
Inconsistency Analysis

- The presence of inconsistency is evident when training models in significantly different subsets of QTs.
- Generalization tends to be less of a challenge in easier operating regions.

Impact over Pipeline Datasets

- The impact is still significant, especially for sources falling in more challenging operating regions ($\mathcal{K}$).

Impact over the Pipeline Datasets



Conclusions

- Even when using diverse datasets, there may be significant inconsistency between unknown QTs and those used in training:
 - Not all operating regions are impacted in the same way.
 - This effect can be substantial in realistic scenarios.
- One should ideally find a way to target the more challenging operating regions.

References

- [1] Q. Giboulot, R. Cogranne, D. Borghys, and P. Bas, "Effects and solutions of cover-source mismatch in image steganalysis," *Signal Process. Image Commun.*, vol. 86, p. 115888, 2020.
- [2] Y. Yousfi and J. Fridrich, "JPEG steganalysis detectors scalable with respect to compression quality," *Electronic Imaging*, vol. 2020, pp. 75–1, 01 2020.
- [3] J.-S. Park, D. Cho, W. Ahn, and H.-K. Lee, "Double JPEG detection in mixed JPEG quality factors using deep convolutional neural network," in *ECCV*, 2018.