Steganalysis in Directional JPEG Images

Benedikt Lorch, Rainer Böhme
University of Innsbruck, Austria
{benedikt.lorch, rainer.boehme}@uibk.ac.at

***Abstract*—A common belief in image steganalysis is that the image orientation does not matter because the statistics are similar when images are scanned horizontally or vertically. However, feature-based and deep learning-based steganalysis is sensitive to image rotations, demonstrating that images do contain directional statistics. In this paper, we systematically study how JPEG steganography and steganalysis methods interact with two relevant causes of directionality: scene content and asymmetries in the JPEG quantization table (QT). We find that (1) steganalysis detectors often achieve higher accuracy in images with directional scene content, (2) asymmetries in the QT bias the detector's generalization towards one direction, (3) a steganalyst can benefit from training a detector tailored to the directionality, (4) rotation augmentation improves orientation robustness, but reduces detection performance on the original orientation.**

***Index Terms*—JPEG steganography, steganalysis, directionality, rotation augmentation**

I. Introduction

Image steganalysis is the task of detecting whether an innocently looking image contains steganography. A common assumption in the steganalysis community is that images are *non-directional*, that is statistics of pixel sequences in horizontal direction are similar to statistics of pixel sequences in vertical direction. This assumption dates back at least to the time of the first quantitative detectors for LSB replacement ("What holds when scanning an image horizontally ought to hold, in general, when scanning vertically" [1, p. 13]). In feature-based steganalysis, the assumption of non-directionality allows feature descriptors to average co-occurrences calculated in horizontal and vertical directions. Popular examples include the subtractive pixel adjacency model (SPAM) ("the effect of portrait/landscape orientation is negligible" [2, p. 3]), the spatial rich model (SRM) [3, p. 5], and the JPEG rich model (JRM) ("assuming the statistics of natural images do not change after mirroring about the main diagonal" [4, p. 3]).

If image statistics were non-directional, then steganalysis detectors trained with images of a single orientation should naturally generalize to other orientations. To test this hypothesis, we train four feature-based detectors for the popular JPEG steganography methods J-UNIWARD [5], UERD [6], J-MiPOD [7], and nsF5 [8], [9] using the 512×512 images from the ALASKA2 dataset compressed with the JPEG quality 75. All four detectors achieve an accuracy between 79.9 % (J-MiPOD) and 94.5 % (nsF5), as shown by the blue bars in

This work is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101021687 (UNCOVER). The numerical results were calculated on the University of Innsbruck's LEO5 cluster and the Vienna Scientific Cluster (VSC).

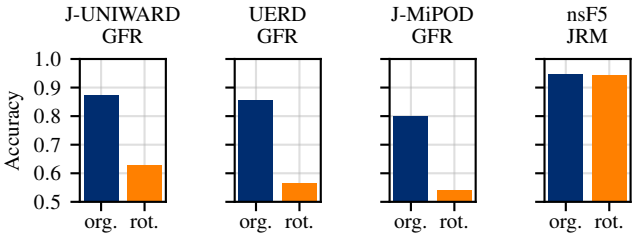


Fig. 1: Motivation: Some feature-based steganalysis methods lose performance when test images are rotated by 90 degrees.

Fig. 1. When the test images are rotated by 90 degrees, the detection accuracy drops by 24.3 %-pt., 29.1 %-pt., and 25.9 %-pt. for J-UNIWARD, UERD, and J-MiPOD, respectively. This demonstrates that cover and stego images contain directional statistics that are exploited by the detectors. An exception is the nsF5 detector, where accuracy drops by only 0.3 %-pt (see Sec. IV for explanation). This shows that steganography and steganalysis interacts in different ways with a directionality that warrant focused study.

Tracing this sensitivity to orientation requires a three-step analysis: we need to understand the directional properties of cover images, how directionality influences the selection channel, and how steganalysis methods deal with directionality. The first step is covered by related work describing the causes of directionality in natural images [10]. Two of the main causes relevant for JPEG steganalysis are the scene content and asymmetries in the JPEG quantization table (QT). However, we still lack a fundamental understanding of how steganography and steganalysis methods contend with directionality.

This paper aims to fill this gap by making the following three contributions. First, we provide a systematic review of how JPEG steganography methods interact with directionality. Second, we investigate orientation sensitivity and symmetrization in hand-crafted features. Third, we empirically analyze several consequences of directionality. We find that steganalysis detectors often achieve higher accuracy in images with directional scene content. Moreover, asymmetries in the QT bias the generalization of the detector towards one direction. Steganalysts may benefit from training detectors specialized to one directionality. By contrast, rotation augmentation reduces orientation sensitivity at the cost of lower detection performance on the original orientation.

The paper is organized as follows: Section II recalls how scene content and QTs introduce directionality. It also introduces our experimental setup and presents baseline results.

Section III analyzes how the selection channel of popular JPEG steganography methods is influenced by directionality. Section IV analyzes how JPEG steganalysis features are affected by directionality. In Section V, we propose ways to deal with directionality. Section VI concludes with a discussion.

II. Background

Directionality can be inherent in the scene content or introduced in several stages during image acquisition [10]. This paper focuses on arguably the two most influential factors in JPEG steganalysis: the scene content and the QT.

a) Scene content: In natural images, most edges align with the horizontal or vertical axes [11]. Many popular image datasets are dominated by images with more horizontal than vertical edges [10]. To quantify directionality in the scene content, we follow [10, long version, Appx. B] and measure the relative distribution of horizontal and vertical frequencies in the image's power spectrum in different frequency bands using steerable pyramids [12]. Analyzing the two highest frequency bands yields a directionality score $d \in [-2, 2]$, where images with negative d contain more horizontal edges and images with positive d contain more vertical edges. Typical example images in ALASKA2 [13] with a strong negative d depict water with horizontal wave fronts. Typical instances with high positive d contain textures of hair, blades of grass, or fences. Exemplary images are shown in [10, Fig. 5]. The directionality score is largely unaffected by JPEG compression, because it integrates over a range of spatial frequencies. This makes it a suitable metric for quantifying directionality in the scene content.

Previous work [10] found that especially the BOSSBase [14] benchmark dataset and to lesser extent also ALASKA2 are skewed towards images with more horizontal edges. Splitting the ALASKA2 dataset into four groups by directionality, with d in $[-2, -1], (-1, 0], (0, 1], (1, 2]$, gives a relative distribution of 17, 44, 33, and 6 %. A training set bias can cause a steganalysis detector to overfit to the directionality that is overrepresented in the training set [10]. To prevent the results from being influenced by such dataset bias, all experiments in this paper use a copy of ALASKA2 in which each image has been rotated by a random multiple of 90 degrees.

b) Quantization table (QT): JPEG compression achieves much of its storage efficiency by quantizing DCT coefficients according to their visual perceptibility. Although camera manufacturers customize their QTs [15], a popular choice are the example tables provided in Annex K of the JPEG standard [16]. Their inclusion in *libjpeg* has made them the de-facto standard in many software libraries and in academic research. Derived from psychovisual thresholding experiments [17, Ch. 5], the luminance QT contains several asymmetric features in all frequency bands. Although these asymmetries were already pointed out in 1992 [18], they have been neglected in the information security and forensics community so far. The vast majority of luminance QTs in forensic datasets are either scaled variants of the standard QT or contain other asymmetric features [10, Tab. 2].

TABLE I: Detection accuracy and sensitivity to rotated images.

			Sym. QT-75		Std. QT-75		Asy. QT-70\80	
Embedding		Feat.	org.	rot.	org.	rot.	org.	rot.
J-UNI.	0.4	GFR	86.9	86.9	87.2	62.9	85.8	53.7
UERD	0.4	GFR	85.5	85.4	85.5	56.4	84.3	51.5
J-MiPOD	0.4	GFR	79.9	79.8	79.9	54.0	79.2	51.1
nsF5	0.2	JRM	94.8	94.8	94.5	94.2	94.7	90.6
J-UNI.	0.4	CNN	90.1	90.1	90.3	77.6	89.1	77.0
UERD	0.4	CNN	92.7	92.7	93.4	87.2	92.3	90.3
J-MiPOD	0.4	CNN	88.9	88.9	88.4	78.8	87.7	82.0
nsF5	0.2	CNN	86.4	86.5	87.0	73.4	86.8	67.6

To study the effect of asymmetries in the QT, we experiment with the following three QTs ordered by increasing asymmetry: a symmetrized version of the standard QT at quality factor (QF) 75 (denoted as *Sym. QT-75*), the standard QT at QF 75 (denoted as *Std. QT-75*), and an artificial asymmetric QT containing the values of the QF-70 QT below the diagonal, of the QF-80 QT above the diagonal, and their average on the diagonal (*Asy. QT-70\80*). We selected this asymmetric QT because it is close to the QF-75 QT in terms of the sum of quantization factors.

c) Experimental setup: All experiments draw from the randomly rotated ALASKA2 color images with size 512^2 [13]. We additionally create three copies rotated by 90, 180, and 270 degrees prior to compression. All images are JPEG-compressed with one of the three QTs and 4:2:0 chroma subsampling using *libjpeg-turbo* 2.1.0. Steganography is embedded into the luminance channel using J-UNIWARD, UERD, J-MiPOD, and nsF5 with an embedding rate of 0.4 bits per non-zero AC coefficient (bpnzAC). For nsF5, we lower the embedding rate to 0.2 because nsF5 introduces more changes.

Our feature-based and CNN-based experiments differ in their experimental setup. Hand-crafted features are extracted from the luminance channel and classified with an ensemble of Fisher linear discriminant (FLD) base learners [19]. The FLD ensembles are trained and evaluated with 10k randomly selected images each, unless otherwise stated. Each image is included twice, once with features from our base and from the 90 degree rotated copy. This ensures that there is no directional bias in the training or evaluation set.

Furthermore, we train EfficientNet-B0 detectors on the decompressed RGB images. We decided not to remove any stride as this would increase GPU memory usage and force us to use smaller batch sizes. In the CNN experiments, images are split into 80 % for training, 10 % for validation, and 10 % for testing. To reduce overfitting, the CNNs are trained with all four database variants and flipping augmentation. All experimental results with FLD ensembles and CNN detectors are averages over ten or five training–test splits, respectively.

d) Baseline results: Prior work suggests that the first three embedding functions are best detected with JPEG phase-aware features in the spatial domain, while nsF5 is best detected in the DCT domain [20], [21], [22], [7]. Hence, the feature-based detectors use the Gabor filter resid-

TABLE II: Summary of sensitivity of embedding functions to directionality in the content and quantization table (QT).

Embedding	Directional content	Directional QT
J-UNIWARD	small	medium
UERD	none	high
J-MiPOD	small	high
nsF5	medium	small

uals (GFR) [22] in the decompressed domain and the JPEG rich models (JRM) [4] in the DCT domain.

The top part of Tab. I reports the test accuracy for feature-based steganalysis. J-MiPOD is the most difficult embedding function to detect, while nsF5 is detected with the highest accuracy, despite the lower embedding rate. With the symmetric QT-75, all detectors generalize to rotated test images. This is expected because the random rotation of the database eliminated any systematic scene directionality. With the standard QT-75, the GFR-based detectors suffer performance drops on rotated test images. (This data is the basis of Fig. 1.) With the asymmetric QT-70\80, the FLD ensemble's performance on rotated images is close to random guessing. The nsF5 detector is only slightly worse when tested with rotated images. We will get back to this and explain it below.

The CNN detectors in the bottom part of Tab. I show the highest accuracy for UERD and the lowest for nsF5. The latter could probably be improved by training with DCT coefficients [23]. As with the ensemble detectors, the performance of the CNNs also drops with rotated test images.

For reference, we also trained SRNet [24] after ImageNet pretraining on the task of detecting J-UNIWARD. With the symmetric QT-75, SRNet achieves an accuracy of $93.4\,\%$ on both original and rotated test images. With the standard QT-75 and the asymmetric QT-70\80, SRNet's accuracy of $93.2\,\%$ and $92.1\,\%$ on the original orientation drops to $83.1\,\%$ and $69.4\,\%$ on rotated test images, respectively. While SRNet outperforms EfficientNet, it takes significantly longer to train. Since no architectural component in SRNet suggests that it would handle directionality differently from EfficientNet, we use the latter for the remainder of the study.

III. Effect of Directionality on Steganography

A prerequisite for studying the influence of steganalysis to directionality is to understand how JPEG steganography interacts with directionality. This section provides a qualitative analysis by inspecting the embedding change probabilities for an embedding rate of 0.4 bpnzAC.

Figure 2 shows how each embedding function reacts to directional scene content and to the three QTs. In each subfigure, the large squares depict the 8×8 DCT subbands from 1 000 randomly selected images, therefore any content directionality should be averaged out. The small squares in the bottom right corners show the same analysis for 1 000 images with directional scene content ($d < -1.5$). These have fewer non-zero AC coefficients in horizontal than in vertical subbands. Brighter colors represent higher embedding probability (lower embedding cost).

J-UNIWARD, UERD, and J-MiPOD have in common that they calculate the embedding cost per DCT coefficient as a ratio of the QT and the image content. The embedding cost scales with the quantization factor in the numerator, as shown by the columns in Fig. 2 a,b,c. The image content in the denominator controls the content adaptivity, so that the embedding prefers textured over smooth regions. Its realization varies between the three embedding functions.

Turning to specific embedding functions, **UERD** measures block flatness, but treats both directions equally. Hence, UERD ignores directionality in the scene content, which can be seen from the similarity of the large and small squares in Fig. 2b. **J-UNIWARD** measures how altering a single cover pixel distorts the horizontal, vertical, and diagonal Wavelet-filtered residuals. If the image content is smoother in one direction than the other, the cost of embedding along orthogonal edges increases. For example, when an image is blurred in the horizontal direction (e.g., motion blur smoothing out vertical edges, $d < 0$), the cost of embedding around vertical edges increases. The small squares in Fig. 2a show a higher embedding probability for vertical subbands compared to horizontal ones, because directional images contain only few or attenuated horizontal frequencies (vertical edges). Yet, the asymmetric QF-70\80 QT overrides the directionality of the image. In **J-MiPOD**, the embedding cost is controlled by the residual pixel variances of the decompressed image. Intuitively, pixels with higher variance are assigned lower embedding cost. In the directional image subset, we observe slightly higher embedding change probabilities in the first and second horizontal subbands compared to their vertical counterparts, Fig. 2c. However, the effect of the scene content is low. Finally, **no-shrinkage F5** (nsF5) does not embed into DC and zero-AC coefficients. Hence, if the image contains more zero-AC coefficients in one direction, nsF5 embeds less into this direction. Directionality in the QT affects nsF5 indirectly as the quantization can zero out AC coefficients, but the effect of the QT is small, as can be seen by comparing the columns in Fig. 2d.

Table II summarizes these findings qualitatively.

IV. Effect of Directionality on Steganalysis

We first focus on feature-based methods, as the effect of directionality is more easily explained with feature descriptions. Results on CNN-based detectors are given in Section V.

Some steganalysis feature sets average submodels calculated in horizontal and vertical directions. Examples are SPAM686 [2], the spatial rich models (SRM) [3], and the JPEG rich models (JRM) [4]. This *symmetrization* lowers the feature dimensionality, thereby reducing the risk of overfitting and multicolinearity in classification tasks. More recent feature sets, such as the DCT residuals (DCTR) [20], the phase-aware rich model (PHARM) [21], and the Gabor filter residuals (GFR) [22], do not average over orientations. While

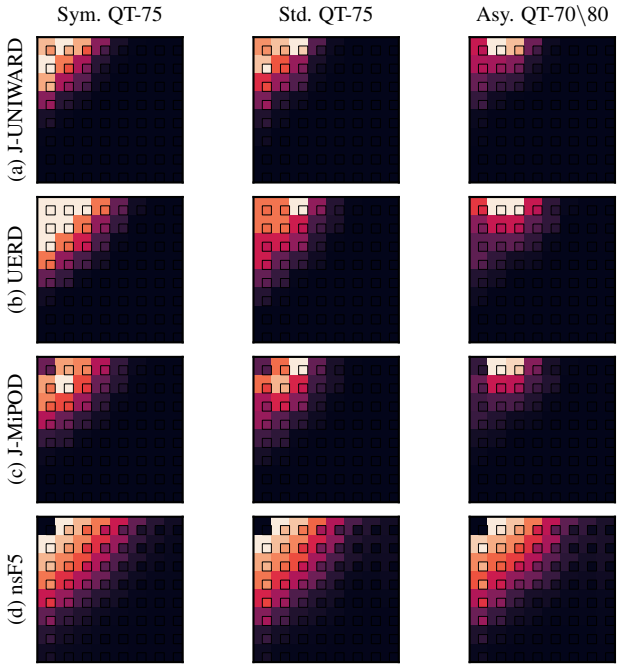


Fig. 2: Interaction of embedding functions with directionality; change probability by DCT subband; balanced images (full squares) and directional images ($d < -1.5$, corner squares). Look for differences in shade between background and corner.

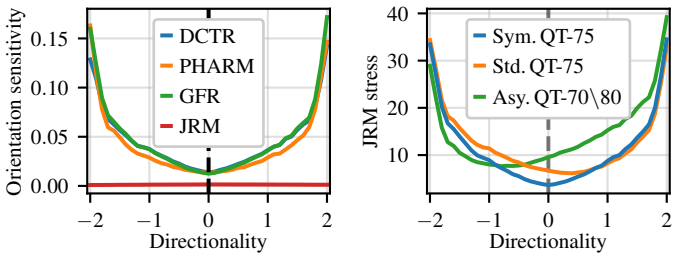


Fig. 3: Left: Comparing features extracted from the original orientation and after rotation by 90 degrees, we observe differences grow with increasing directionality. JRM features seem insensitive. Right: Feature symmetrization causes stress *inside* the JRM features. Asymmetric QTs skew the distribution.

PHARM and GFR average by flipping, we did not find any explanation why directional symmetrization was abandoned.

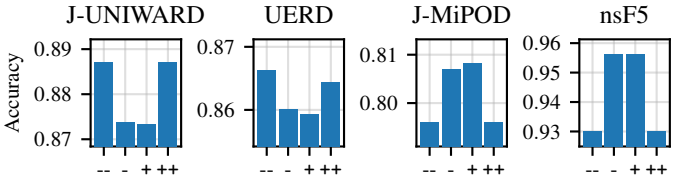
a) Orientation sensitivity of phase-aware features: We evaluate the orientation sensitivity of the DCTR, PHARM, and GFR features by comparing feature descriptors extracted from the same image in its original orientation and rotated (after compression) by 90 degrees. To account for the varying number of submodels with different sizes, we normalize each submodel to unit sum, compute the L1 distance between the submodel from original and rotated image, and average these distances across submodels. Figure 3 (left) shows the mean L1 distance split by directionality scores from our randomly rotated ALASKA base compressed with the symmetric QT-75. The descriptors are closest when the scene content has no dominant direction ($d \approx 0$), but differences increase with stronger directionality, demonstrating that these features capture directional statistics.

b) Orientation symmetrization in the JPEG rich model: JRM remains the most relevant orientation-symmetrized feature descriptor for JPEG steganalysis, as the SRM has been superseded by the (orientation-sensitive) PHARM descriptor [21]. Figure 3 (left) suggests that the (symmetrized) JRM features are barely sensitive to image rotation. While orientation symmetrization reduces orientation sensitivity at the outside, averaging horizontal and vertical submodels creates *stress* inside the feature descriptor. To measure stress, we disable the directional symmetrization and compute the (unnormalized) L1 distances between original and the transposed submodels, which the JRM usually symmetrizes.

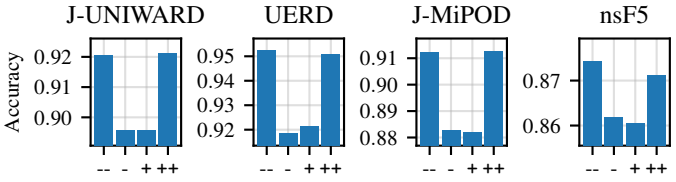
Figure 3 (right) relates the JRM stress to the image directionality. With the symmetric QT-75, images with $d \approx 0$ cause the lowest stress. The stress increases with the directionality of the scene content. With the standard QT-75, the lowest stress is observed for images with a directionality score around $d \approx 0.4$. This is because the standard QT-75 contains lower quantization factors in the vertical mid-frequencies than in the horizontal mid-frequencies. Images with small positive directionality (more horizontal frequencies) compensate for the slight asymmetry in the QT. The artificially asymmetric QT-70\80 shifts the lowest stress to images with $d \approx -0.8$. This QT has lower quantization factors for the horizontal DCT coefficients and higher quantization factors for the vertical coefficients. Images with moderately negative d compensate for the asymmetric quantization. With this asymmetric QT, the DCTR and GFR features exhibit a similar trend, i.e., lower orientation sensitivity for images with small negative directionality, although the trend is less pronounced than for the JRM features. (Plot omitted due to space constraints.)

While the averaged curves in Fig. 3 (right) suggest a clear connection between directionality and stress, we also observe individual images with high directionality but low stress. These images contain mostly smooth content. The images with the highest stress show scenes with high-frequency patterns. Future work could tune the directionality detector to only consider frequency bands that are captured by the JRM features.

c) Are the JRM features invariant to rotation?: Despite the symmetrization, JRM features are not invariant to rotation, as shown by the steganalysis results with nsF5 and the asymmetric QTs in Tab. I. One reason is that submodels extracted along the major diagonal correspond to submodels extracted along the minor diagonal in rotated images. Hence, the submodels switch places in the flattened feature descriptor. Additionally, the JRM applies only sign symmetrization to the two-dimensional co-occurrences, but no directional symmetrization like in the SRM [3, Sec. II,C]. Analyzing nsF5 with the Std. QT-75 and GFR features yields an accuracy of 84.8 % and a performance drop by 23.8 %-pt. on rotated images. This indicates that the high degree of robustness still originates from the JRM features and less from nsF5.



(a) Features + FLD ensemble detectors



(b) CNN detectors

Fig. 4: Grouping the test images by directionality shows that all CNN and the feature-based J-UNIWARD and UERD detectors are more accurate on images with strong directionality.

V. Dealing with Directionality

The baseline results in Tab. I demonstrate that steganalysis detectors are sensitive to rotated images with asymmetric QTs. Since the QT is available to the analyst, performance losses due to asymmetric QTs can be mitigated when steganalysts are aware of the detectors' limitations and carefully control image rotations applied during image decompression.

a) J-UNIWARD and UERD detectors achieve higher accuracy in directional images: Having trained feature-based and CNN-based detectors in Sec. II, we report separately the accuracy for images of different directionality with the symmetric QT-75. The test images are split by directionality into four groups $d \in [-2,-1], (-1,0], (0,+1], (+1,+2]$. For brevity, the groups are denoted as $--$, $-$, $+$, and $++$. Fig. 4 shows the feature-based FLD (top) and CNN detectors (bottom). All CNNs and the FLD detectors for J-UNIWARD and UERD achieve higher test accuracy on images with stronger directionality ($--$ and $++$). Only the FLD detectors for J-MiPOD and nsF5 achieve higher accuracy on less directional images ($-$ and $+$). The discrepancy between FLD and CNN detectors for J-MiPOD and nsF5 suggests that steganography is not inherently less secure in directional images. Instead, the detection accuracy across directionality groups also depends on which embedding artifacts are exploited by detector. Analyzing the detector could reveal if accuracy differences stem from the steganography method or learned artifacts, but this is beyond this paper's scope.

b) When the QT is asymmetric, detectors generalize better to one direction: The previous experiment showed that detectors trained with a symmetric QT generalize, on average, equally well to images of negative and positive directionality. We now show that asymmetric QTs can bias the detector to generalize better to images of one direction. Figure 5 displays the test accuracy of three J-UNIWARD GFR-FLD detectors, one for each QT, split by image directionality. As before, the symmetric QT makes the detector generalize equally well to

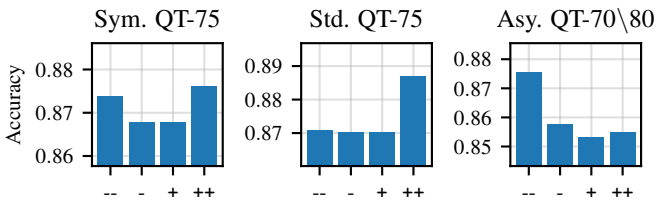


Fig. 5: Asymmetric QTs bias the generalization towards one direction: With the standard QT, detectors work better on image with positive directionality. With the asymmetric QT, detectors work better on images with negative directionality.

TABLE III: Rotation augmentation makes CNN detectors generalize to rotated images, but they underperform CNNs trained with a specific compression direction.

		Base rotation		Rotation aug.			
Embedding and rate		org.	rot.	org.	rot.	Δ org.	Δ rot.
J-UNIWARD	0.4	90.3	77.6	89.4	89.3	−0.9	+11.7
UERD	0.4	93.4	87.2	91.9	91.8	−1.5	+4.6
J-MiPOD	0.4	88.4	78.8	84.7	84.8	−3.7	+6.0
nsF5	0.2	87.0	73.4	85.7	85.7	−1.3	+12.3

images of negative and positive directionality. With the standard QT, the detector generalizes better to images of positive directionality. With the asymmetric QT-70\80, the detector generalizes better to images of negative directionality. This corroborates the observation from Sec. IV, where QT asymmetries skew the orientation sensitivity towards the content direction that compensates for the asymmetric quantization. The two groups with strong directionality are still detected with higher accuracy, because the effect of the asymmetric QT overlaps with the observation of higher detectability in strongly directional images described above. We observe the same trend for UERD and J-MiPOD detectors, but not for nsF5. Presumably nsF5 is less sensitive to QT asymmetries.

c) Analyzing directional images can benefit from a specialized detector: In practice, the steganalyst may find suspicious images with directional content. Suppose the test images are dominated by vertical edges ($d > 1$, group $++$) and use the symmetric QT-75. We show that it can be beneficial to train a specialized detector for this directionality. We randomly split the images with $d > 1$ into two halves for training and testing (4528 images each), and train ensemble classifiers as before. The results are averaged over ten training–test splits. The orange bars in Fig. 6 show the test accuracy on the directional data. The proposed specialized detector outperforms a general detector trained with the same number of images but randomly selected from all directionalities (dark blue). For reference, a detector that is trained with four times the number of randomly selected images (light blue) achieves about the same accuracy as the specialized detector. When training material is abundant, the steganalyst does not benefit from a specialized detector. However, when steganalysts need to collect a training database for a specific case or channel, they can benefit from matching the directionality of the suspicious images.

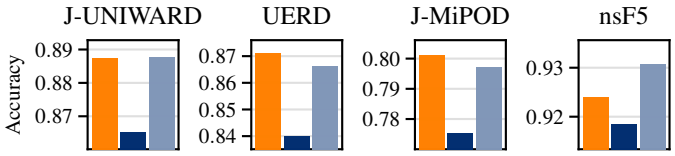


Fig. 6: When test images are directional ($d \in [1, 2]$), a specialized detector trained with images of similar directionality (orange) outperforms a detector trained with the same number of randomly selected images (dark blue). Increasing the number of randomly selected training images by factor 4 compensates for the lack of specialization (light blue).

d) Training CNN detectors with rotation augmentation: To make CNNs robust to all orientations, an intuitive approach is to randomly rotate the training images after decompression, as is common practice [25]. We retrain CNN detectors with rotation augmentation on the base dataset with the standard QT-75 and compare with the results from Tab. I, where images were rotated prior to compression (*base rotation*). The results in Tab. III show that rotation augmentation allows CNNs to generalize to rotated test images. At the same time, rotation augmentation decreases accuracy on the original orientation by 0.9 %-pt. (J-UNIWARD) to 3.7 %-pt. (J-MiPOD).

VI. Concluding Discussion

We have analyzed the effects of directionality on steganography embedding functions and detection performance. Our experiments are broad, but far from comprehensive. We have focused on two factors of directionality, the scene content and the QT. All images were compressed using the Loeffler–Ligtenberg–Moschytz 2-D DCT (*ISLOW*) and with 4:2:0 chroma subsampling, because these are *libjpeg*'s default settings and most likely encountered in practice. Both operations produce directional rounding artifacts [10], but we deemed their effect negligible due to the subsequent quantization. Such rounding artifacts may be more pronounced at higher QFs.

Even in directionality-balanced setups, e.g., randomly rotated dataset and symmetric QT, we observed that individual detectors became biased towards one direction. In FLD ensemble classifiers, this could be because individual base learners are trained on random subsets of training samples, which might by chance be biased. In the CNNs, this could be related to ImageNet pre-training, but further experiments are needed to pinpoint the exact cause. We conjecture that rotation augmentation promotes learning of orientation-invariant features, but discards useful directional image properties.

Future work could investigate if the improvements in Sec. V-c) and d) can be achieved with fine-tuning rather than retraining. Our directionality score derived from steerable pyramids integrates over a large range of spatial frequencies. Tuning the directionality score to consider only the frequency bands that are relevant to steganography and steganalysis could provide an even finer analysis of the effects of directionality. Broader future work should also investigate the effect of directionality in spatial domain steganography.

References

[1] A. Ker, "Quantitative evaluation of pairs and RS steganalysis," in *Security, Steganography, and Watermarking of Multimedia Contents*, vol. 5306, 2004, pp. 83–97.

[2] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215–224, 2010.

[3] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[4] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," in *Media Watermarking, Sec., and Forensics*, vol. 8303, 2012.

[5] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, p. 1, 2014.

[6] L. Guo, J. Ni, W. Su, C. Tang, and Y. Shi, "Using statistical image model for JPEG steganography: Uniform embedding revisited," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2669–2680, 2015.

[7] R. Cogranne, Q. Giboulot, and P. Bas, "Efficient steganography in JPEG images by minimizing performance of optimal detector," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1328–1343, 2022.

[8] A. Westfeld, "F5 — A steganographic algorithm," in *Information Hiding*, vol. 2137, 2001, pp. 289–302.

[9] J. Fridrich, T. Pevný, and J. Kodovský, "Statistically undetectable JPEG steganography: Dead ends challenges, and opportunities," in *ACM Workshop on Multimedia Security*, 2007, pp. 3–14.

[10] B. Lorch and R. Böhme, "Landscape more secure than portrait? Zooming into the directionality of digital images with security implications," in *USENIX Security Symposium*, 2024, pp. 6903–6920.

[11] D. Coppola, H. Purves, A. McCoy, and D. Purves, "The distribution of oriented contours in the real world," *National Academy of Sciences*, vol. 95, no. 7, pp. 4002–4006, 1998.

[12] E. Simoncelli and W. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *IEEE International Conference on Image Processing*, vol. 3, 1995, pp. 444–447.

[13] R. Cogranne, Q. Giboulot, and P. Bas, "ALASKA#2: Challenging academic research on steganalysis with realistic images," in *IEEE Workshop on Information Forensics and Security*, 2020.

[14] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system: The ins and outs of organizing BOSS," in *Information Hiding*, vol. 6958, 2011, pp. 59–70.

[15] E. Kee, M. K. Johnson, and H. Farid, "Digital image authentication from JPEG headers," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1066–1075, 2011.

[16] International Telecommunication Union, "ISO/IEC 10918-1: 1993(E) CCIT Recommendation T.81," w3.org/Graphics/JPEG/itu-t81.pdf, 1993.

[17] H. Lohscheller, "Einzelbildübertragung mit wachsender Auflösung," Dissertation, Technische Hochschule Aachen, 1982.

[18] S. Klein, D. Silverstein, and T. Carney, "Relevance of human vision to JPEG-DCT compression," in *Human Vision, Visual Processing, and Digital Display III*, vol. 1666, 1992, pp. 200–215.

[19] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2012.

[20] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

[21] ——, "Phase-aware projection model for steganalysis of JPEG images," in *Media Watermarking, Security, and Forensics*, vol. 9409, 2015.

[22] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of adaptive JPEG steganography using 2D Gabor filters," in *ACM Workshop on Information Hiding and Multimedia Security*, 2015, pp. 15–23.

[23] Y. Yousfi and J. Fridrich, "An intriguing struggle of CNNs in JPEG steganalysis and the OneHot solution," *IEEE Signal Processing Letters*, vol. 27, pp. 830–834, 2020.

[24] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2019.

[25] T. Itzhaki, Y. Yousfi, and J. Fridrich, "Data augmentation for JPEG steganalysis," in *IEEE Workshop on Information Forensics and Security*, 2021.