

An Empirical Study of Steganography and Steganalysis of Color Images in the JPEG Domain

Théo Taburet¹, Louis Filstroff³, Patrick Bas¹, Wadih Sawaya²

¹ Univ. Lille, CNRS, Centrale Lille, UMR 9189
CRIStAL, F-59000 Lille, France

² IMT Lille-Douais, Univ. Lille, CNRS, Centrale Lille, UMR 9189
CRIStAL, F-59000 Lille, France

³ IRIT, Université de Toulouse, CNRS, France

Abstract. This paper tackles the problem of JPEG steganography and steganalysis for color images, a problem that has rarely been studied so far and which deserves more attention. After focusing on the 4:4:4 sampling strategy, we propose to modify for each channel the embedding rate of J-UNIWARD and UERD steganographic schemes in order to arbitrarily spread the payload between the luminance and the chrominance components while keeping a constant message size for the different strategies. We also compare our spreading payload strategy w.r.t. two strategies: (i) the concatenation of the cost map (CONC) or (ii) equal embedding rates (EER) among channels. We then select good candidates within the feature sets designed either for JPEG or color steganography. Our conclusions are threefold: (i) the GFR or DCTR features sets, concatenated on the three channels offer better performance than ColorSRMQ1 for JPEG Quality Factor (QF) of 75 and 95 but ColorSRMQ1 is more sensitive for QF=100, (ii) the CONC or EER strategies are suboptimal, and (iii) depending of the quality factors and the embedding schemes, the empirical security is maximized when between 33% (QF=100, UERD) and 95% (QF=75, J-UNIWARD) of the payload is allocated to the luminance channel.

1 Introduction

Since image steganography may be used to hide potentially sensitive messages inside mainstream image formats, it is surprising to notice that the majority of academic contributions in steganography and steganalysis deals with exotic image formats such as lossless raw coding (PGM, PPM) or grayscale JPEG images. Moreover, if a steganographic implementation addresses the most popular image format of the Web, i.e. color JPEG images, it is usually done without distinguishing the color components.

More accurately, whenever embedding is realized on color numerical images in the pixel domain (usually for steganalysis purposes), it is most of the time implemented independently on each component [1,2,3] but more advanced schemes use a synchronization strategy to have more coherent embedding changes across the different channels [4]. Popular implementations of color JPEG steganography such as F5 [5] or J-UNIWARD [6] alter DCT coefficients without taking into account their related color channels or the related component statistics.

Regarding the steganalysis of color images, A.D. Ker et al. underlined in [7] that most of the research carried out over the past ten years focused on grayscale images, and that methods taking into account correlation between channels were left to be desired. Regarding spatial steganography and *RGB* components, Goljan et al. developed the Spatial and Color Rich Models (SCRMQ1) [3], which can be seen as a spatial extension of the Spatial Rich Model (SRM) [8] that uses Color Co-occurrence Matrices. These features, as well as other specific feature sets, such as the Gaussian filter bank features [1] (which is an extension of the

SCRMQ1), or the CFA-aware features [2], have been particularly designed for color pixel-based steganography. Recently a steganalysis scheme using deep convolutional networks [9] has been proposed, it uses 3 disjoint layers (one for each color components) that are pooled together in the next layers. However the complexity learning phase makes this approach prohibitive to benchmark a large variety of different embedding strategies.

To the best of our knowledge there is no color-specific steganalysis method for JPEG images. For grayscale JPEG images however, among the most advanced feature sets are the DCTR (DCT residuals) [10] and the GFR (Gabor Filter residuals) [11]. The DCTR extraction method computes residuals from convolutions with 8x8 DCT basis vectors and the features are generated by computing histograms on each residual. The same methodology is adopted for GFR by using oriented Gabor kernels instead.

The overall lack of solutions for color JPEG steganography and steganalysis can also be explained by the diversity of color JPEG images that are not all coded in a unique way. For example, since the chroma sub-sampling option is also variable between images (see Section 2), the dimensions of the color components of a JPEG image are dependent on the acquisition device or the developing software.

In this paper we propose to extend the popular J-UNIWARD [6] and UERD [12] algorithms to color JPEG, and to evaluate their detectability by designing appropriate feature sets. The next section explains how to obtain different versions of the embedding scheme by spreading the payload among the color components. Section 3 presents feature sets used in this framework, they are derived from popular feature sets in the literature (SCRMQ1, DCTR and GFR). Section 4 provides the different results and associated conclusions related to the paper, both on the best embedding strategy and the most sensitive feature sets for steganalysis.

2 Practical Optimization of the Embedding Schemes

Without loss of generality, we have decided to study here the 4:4:4 JPEG sampling format. This choice is motivated by the fact that, as presented in Table 1, certain sources such as the photo sharing website Flickr mainly use this sampling strategy. However, note that the proposed methodology (choice of the spreading factor in Section 2, choice of the most sensitive feature sets in Section 3), can also be adopted for other sampling formats.

Chroma sub-sampling	4:4:4	4:2:2	4:2:0
Proportion	65%	8%	27%

Table 1: Statistics of chroma sampling strategies for 10,000 "Explored" images downloaded at full resolution from Flickr.com.

Furthermore, because of their excellent performances, we decided to adapt J-UNIWARD and UERD algorithms, respectively proposed by Holub et al and Guo et al. J-UNIWARD is a ternary adaptive embedding scheme which computes costs for each DCT coefficients based on the impact of a ± 1 modification on the wavelet decomposition of the spatial representation of the image. UERD uses a different approach to compute the cost, which is based on the DCT coefficients variation within a block and its neighboring blocks.

2.1 Parametrization of the Payload Distribution for YC_bC_r Components

Because a JPEG color image is composed of 3 color components, it is not straightforward to spread the payload among Y , C_b and C_r . Note that in order to provide a fair comparison, the same message size has to be embedded for each spreading strategy. This problem can be seen as a problem of batch steganography [13], and hence we can use several strategies to allocate the total payload within the three color components. We list below 3 natural ways to deal with this issue:

- **Cost map concatenation (CONC):** a first strategy is to compute a common cost map by firstly concatenating the YC_bC_r components, secondly computing a joint distortion map, and finally computing the embedding probability for the embedding rate α .
- **Equal embedding rate (EER):** one straightforward way to perform the embedding is to set the same payload rate for the three color channels, but this strategy omits the fact that the chroma components contain on average less information than the luminance component and that it is quantized in a different way. Table 2 shows average number of non-zero-AC (nzAC) coefficients for BOSSBase in the JPEG 4:4:4 domain. Firstly, we can observe that the number of luminance nzAC coefficients represents between around 36% (QF=100) and 80% (QF=75) of total number of nzAC coefficients. Secondly, since chroma components are less informative, for an equal embedding rate the embedding should be more detectable for the chroma channels. This rational which will be practically assessed in Section 4, motivates the following more flexible strategy.
- **Arbitrary repartition of the payload between the 3 channels (ARB):** this strategy, which is detailed in the rest of this section, consists in using a new embedding parameter to arbitrary spread the payload across the YC_bC_r channels.

Component	mean (QF=75)	Ratio	mean (QF=95)	Ratio	mean (QF=100)	Ratio
Y	41340	79%	96893	62%	183715	36%
C_b	6087	11%	29284	19%	157970	31%
C_r	5001	10%	29105	19%	167419	33%

Table 2: Statistics (average number of nzAC coefficients) of BOSSBase for quality factor of 75, 95 and 100, 4:4:4 chroma sampling. BOSSBase in color JPEG was generated by first exporting to PPM format using the standard BOSSBase conversion routine.

We propose to distribute the payload among luminance and chrominance components in the following way. Given N_Y , N_{C_b} and N_{C_r} the number of non-zero AC coefficients (nzAC) for respectively Y , C_b and C_r , and α the total embedding rate per nzAC for the three channels, we set:

- P : the message size, i.e. the payload, in bits, which has to stay constant for all strategies,
- α : the total embedding rate, in bit per nzAC coefficient, as it is classically defined in steganography
- a couple of parameters (β, γ) such that $\gamma(1 - \beta)$ defines the embedding rate associated to the luminance channel (in bit per nzAC luma coefficient) and $\gamma\beta$ the embedding rate associated to the two chrominance channels (in bit per nzAC chroma coefficient). Note that $\beta = 0$ implies that all the payload is embedded in the luminance channel ($\gamma = \alpha$), and $\beta = 1$ means that the whole payload is carried by the chroma channels. γ can be seen as a necessary degree of freedom used to choose $\beta \in [0, 1]$ and to be able to compare embeddings at equal message sizes.

The embedded message size (in bits) is then given by:

$$P = \gamma [(1 - \beta) N_Y + \beta (N_{C_b} + N_{C_r})]. \quad (1)$$

The embedding rate per nzAC α is then:

$$\alpha = \frac{P}{N_Y + N_{C_b} + N_{C_r}} = \frac{\gamma [(1 - \beta) N_Y + \beta (N_{C_b} + N_{C_r})]}{N_Y + N_{C_b} + N_{C_r}}. \quad (2)$$

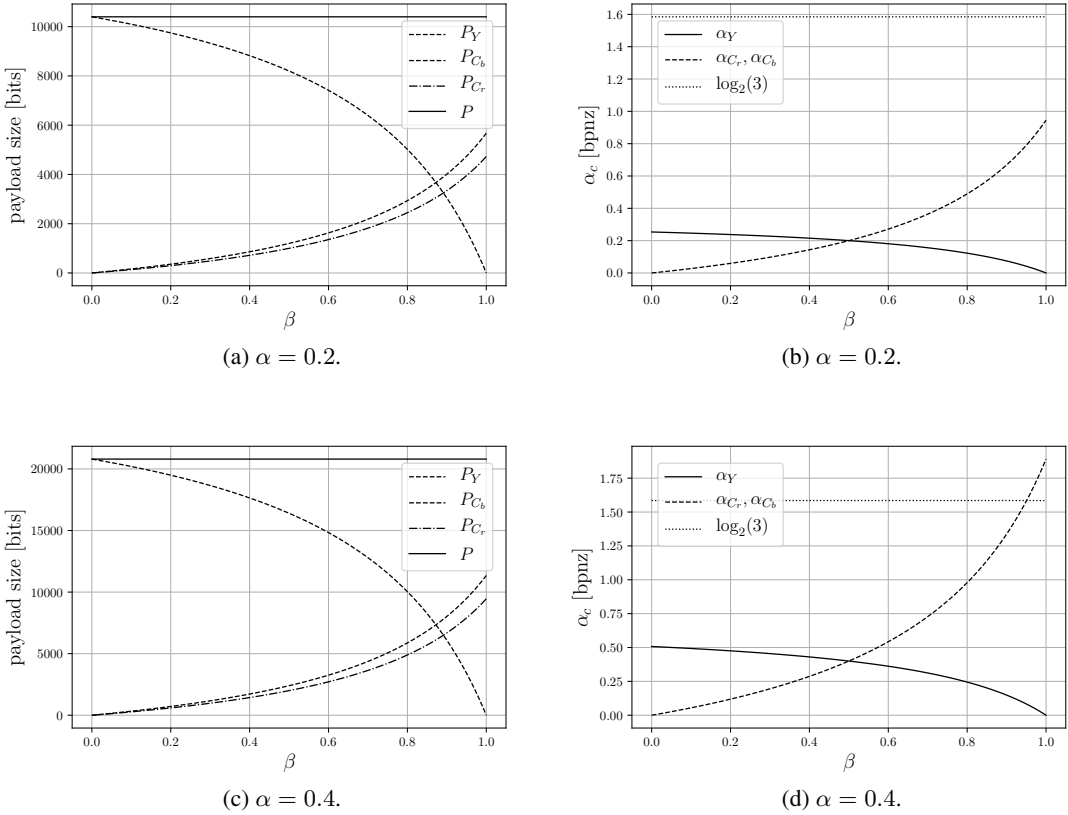


Fig. 1: Left: allocations of the message sizes among the different channels. Right: values of the different embedding rates w.r.t. the parameter β . For this example, the number of nzAC coefficient is close to the one obtained for QF=75: $N_Y = 41000$, $N_{C_b} = 6000$, $N_{C_r} = 5000$.

We consequently have four parameters (α, β, P, γ) and one degree of freedom to choose the embedding rate or the payload. We can set the message size P and choose β , and α and γ will be calculated using (2) and (1) respectively. In a more conventional way, we can set the embedding rate α and choose β , and P and

γ will be computed using (2) and (1) respectively. If we worked on grayscale images, we would set $\beta = 0$ which means $P = \alpha N_Y$.

Note also that the equal embedding rate strategy is equivalent to have $\beta = 0.5$ since in this case all embedding rates are equal to α .

Moreover, the proportion of the payload R_L carried by the luminance channel is given by:

$$R_L = \frac{(1 - \beta) N_Y}{(1 - \beta) N_Y + \beta (N_{C_b} + N_{C_r})}. \tag{3}$$

Fig. 1 illustrates evolutions of message sizes and embedding rates for the three components w.r.t. the parameter β for two embedding rates of 0.2 and 0.4 bit per nzAC and for arbitrary (N_Y, N_{C_b}, N_{C_r}) . We can notice that the embedding rate can be larger than the maximum embedding rate for J-UNIWARD and UERD ($\log_2(3)$ bits), but this happens only for low quality factors, high embedding rates and β close to 1. In Section 4 we shall see that these configurations provide very low practical security and will consequently never be used in practice.

3 Feature Sets for Steganalysis of Color JPEGs

We decide to first benchmark the steganographic scheme by adapting methods dedicated either for color spatial or JPEG grayscale steganographic schemes.

Our first choice is the Color Rich Model, which is composed of the *SRMQ1* features, augmented by a collection of 3D co-occurrences of residuals between color channels to obtain a set of 18,157 features referred in [3] as *SCRMQ1*. Since the images we analyze are embedded in the JPEG domain, we propose an alternative version of the *SCRMQ1* where residuals are computed in the YC_bC_r color space. That means that the *RGB* components are first converted into YC_bC_r and then all residual filtering and co-occurrence computation are performed in the exact same way as for *SCRMQ1*.

Table 3 shows for UERD and J-UNIWARD the difference between computing the *SCRMQ1* in the *RGB* domain or in the YC_bC_r domain, one can see on this example (that generalizes to other embedding rates) the necessity of computing features in the appropriate subspace. This can be explained by the fact the spatial discrepancies captured by the *SRMQ1* features are more significant when applied in the same color space than the embedding. Since no synchronization strategy is applied between the color components, the co-occurrences between the color channels are more effective as well.

Color space	<i>RGB</i>	YC_bC_r
P_E (UERD)	14.38%	8.13%
P_E (J-UNIWARD)	32.39%	8.02%

Table 3: Impact of computing the *SCRMQ1* feature set in the appropriate domain for Color-JPEG steganography. $\alpha = 0.4$, $\beta = 0.7$, QF=95.

Because both DCTR and GFR feature sets provide excellent performance on grayscale images, we also decide to use these features on color images by simply concatenating the features computed for each channel. Here, by definition the features are computed directly in the YC_bC_r color space (only the inverse DCT transform is computed before filtering with DCT kernels for C-DCTR or Gabor kernels for C-GFR), and we end up with $3 \times 8000 = 24,000$ features for Color-DCTR features (abbreviated C-DCTR) and $3 \times 17000 =$

51,000 for Color-GFR (abbreviated C-GFR). Note that even if C-DCTR and C-GFR do not compute co-occurrence matrices between color channels as SCRMQ1, discrepancies between embedding strategies between different channels can be captured by the concatenation operation. For example one can expect that by setting the embedding parameter $\beta = 0$, C-DCTR and C-GFR feature sets will capture a discrepancy between the statistical properties of the luminance channel w.r.t the chroma channels.

4 Results

4.1 Experimental protocol

In order to evaluate the proposed scheme we generate a version of BOSSBase [14] for the available RAW images and we changed the script by directly generating a JPEG image from the cropped and scaled 512x512 PPM image. For all these experiments we choose:

- three JPEG quality factors, 75, 95 and 100,
- two embedding rates, $\alpha = 0.2$ and $\alpha = 0.4$ bit per nzAC coefficient.
- two steganographic schemes J-UNIWARD and UERD (see Section 2)
- the parameter β fluctuating in the range $[0,1]$ to assess the impact of payload allocation, recalling that $\beta = 0.5$ is tantamount to the Equal Embedding Rate strategy. For $\alpha = 0.4$, we do not benchmark the scheme for $\beta = 0.9$ or $\beta = 1.0$ since it can be that in this case the embedding rate in the chroma components is larger than $\log_2(3)$ bits.
- for comparison purposes the CONC strategy is also benchmarked.

All detectors are trained as binary classifiers implemented using the FLD ensemble [15], with default settings. The ensemble by default minimizes the total classification error probability under equal priors:

$$P_E = \min_{P_{FA}} \frac{1}{2}(P_{FA} + P_{MD}),$$

where P_{FA} and P_{MD} denote respectively the false-alarm and missed-detection probabilities. P_E is averaged over ten different training and testing sets, in which the 10,000 cover images and the associated 10,000 stego images are randomly divided into two equal halves for pair-training and testing. We report this value as \bar{P}_E for values of γ satisfying (1) and (2).

4.2 Comparison between Embedding Strategies

We look at the embedding strategies chosen between ARB, CONC and EER (i.e. $\beta = 0.5$) that gives the highest P_E considering the most efficient feature sets, i.e. the minimum of P_E over the 3 feature sets. From these results, different comparisons can be established:

- As a general conclusion, for all feature sets the arbitrary spreading of the payload can allow to achieve the highest practical security.
- For QF=75 (see Figures 2 and 3), it is reached for $\beta \simeq 0.2$ for J-UNIWARD and $\beta \simeq 0.3$ for UERD. For example for J-UNIWARD, using equations (3) and (2), it means that on average 94% of the payload is carried by the luminance channel which itself carries 79% of the nzAC coefficients.
- For QF=95 (see Figures 4 and 5), gives the same conclusions w.r.t. the optimal values of β . In this case however, 85% of the payload is carried by the luminance channel for J-UNIWARD, which conveys 62% of the nzAC coefficients.

- For QF=100 (see Figures 6 and 7), the maximal empirical security is reached for $\beta \simeq 0.3$ for J-UNIWARD and $\beta \simeq 0.4$ for UERD. In this case 50% of the payload is carried by the luminance channel on average for J-UNIWARD and 33% for UERD.

Moreover, two important remarks can be drawn from these extensive sets of results:

- We can see that the naive strategies of setting $\beta = 0$ (all the payload is embedded in the luminance channel) or $\beta = 0.5$ (the embedding rates are equal) are suboptimal. For example for J-UNIWARD at $\alpha = 0.2$ and QF=75 using the best feature set (C-GFR), the gap between $\beta = 0.0$ and $\beta = 0.2$ (the optimal strategy) $\Delta P_E \simeq 2\%$, between $\beta = 0.2$ and $\beta = 0.5$ is $\Delta P_E \simeq 6\%$. These two conclusions are not surprising: pushing all the payload in the luminance channel is equivalent to not taking into account possible dependencies between luminance and chroma components. Furthermore this leads to a concentration of changes on the same component, hence a higher detectability. On the other hand, the EER strategy would be optimal only if the capacity of the scheme would be directly proportional to the number of nzAC, and we know that it is not true in practice (see for example the Ker laws [16]).
- The CONC strategy (concatenation of the costs then embedding), represented by the vertical lines on the different plots, is also clearly sub-optimal for the different embedding rates, embedding schemes or quality factors. For example for J-UNIWARD at $\alpha = 0.2$ and QF=75, the gap between $\beta = 0.2$ and CONC is $\Delta P_E \simeq 13\%$. This can be explained by (i) the fact that empirical costs computed by J-UNIWARD and UERD have been designed w.r.t. grayscale image steganography and they do not take into account potential dependencies between color channels and (ii) the fact that luma and chroma components do not use the same quantization matrices except for QF=100. The mixing between costs computed using completely different quantization steps can explain the non-adaptivity of the CONC strategy. When $QF \neq 100$ we can see that the CONC strategy is however closer to the optimal solution for UERD than for J-UNIWARD.

As a more general comparison, we can see that as for grayscale JPEG steganography the practical security of UERD is slightly more important than the practical security of J-UNIWARD, especially for low embedding rates. For example at QF=95 using C-GFR, the gap is $\Delta P_E \simeq 3\%$ for $\alpha = 0.2$, at QF=100, the gap is $\Delta P_E \simeq 4\%$ for $\alpha = 0.2$.

4.3 Comparison between Feature Sets

We now draw few conclusions on the steganalysis side, depending on the JPEG QF.

For QF=75 and QF=95: the C-GFR feature sets outperforms the C-DCTR feature sets by a small margin for UERD ($\Delta P_E \simeq 3\%$ at QF=75, $\Delta P_E \simeq 5\%$ at QF=95 for optimal β), and the SCRMQ1 features set by a large margin.

C-DCTR shows superior performance w.r.t. the two other feature sets except for J-UNIWARD at QF=95 and $\beta \geq 0.3$ where C-DCTR is more performant. However for the embedding values of β offering the best empirical security, C-GFR are rather efficient.

For QF=100: SCRMQ1 features are more sensitive than the two other feature sets. This can be explained by the fact that lot of information from the uncompressed image is kept at QF=100 since all quantization steps are equal to 1. Consequently, SCRMQ1 computed in the YC_bC_r appears to be one ideal candidate for accurate steganalysis.

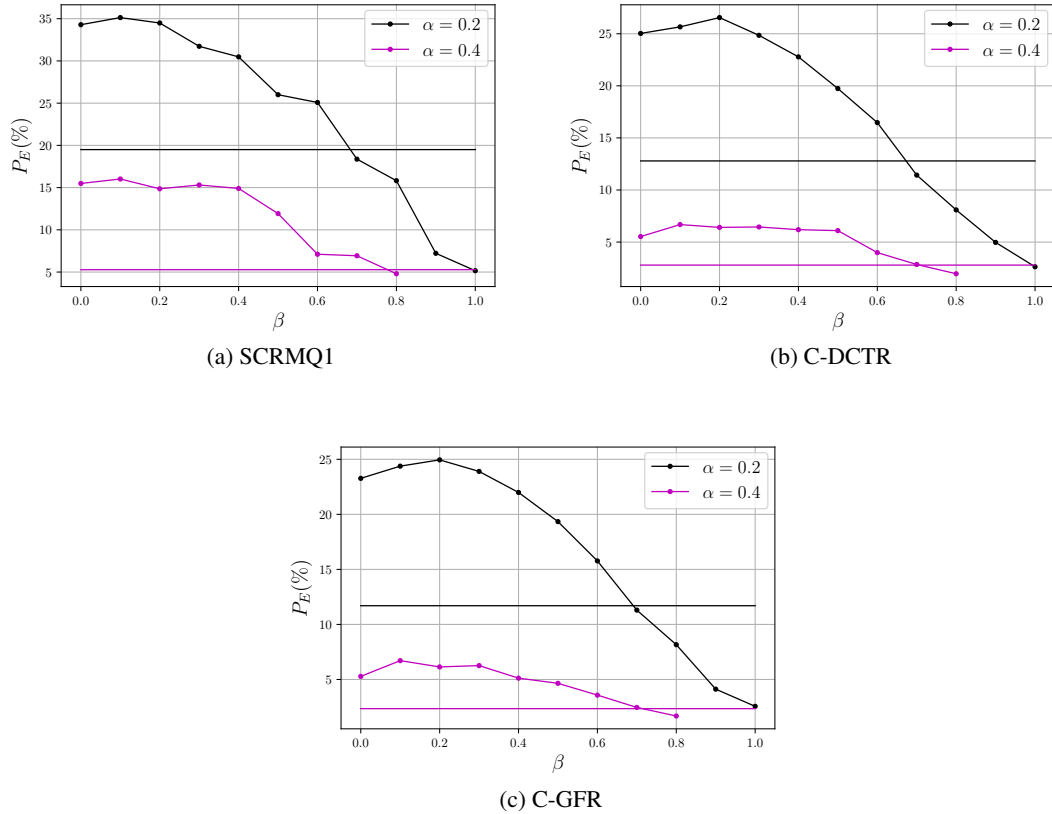


Fig. 2: J-UNIWARD: Comparison w.r.t. β for different feature sets, JPEG QF = 75. Horizontal lines are results for the CONC strategy.

5 Conclusion and perspectives

This paper has proposed an empirical analysis of JPEG steganography and steganalysis on color images. Our conclusions are three-fold: (i) using constant embedding rate across channel or concatenating the cost maps are not optimal embedding strategies since they do not take into account statistical dependencies between the color channels, (ii) especially for JPEG QF of 75 and 95, most of the payload should be concentrated in the luminance channel to maximize empirical security, (iii) over the three reputed feature sets used in color or JPEG steganalysis, the concatenation of GFR features on the 3 channels offer on average the best performance for QF=75 and QF=95, but SCRMQ1 computed in the $Y C_b C_r$ domain offers superior performances for QF=100. Future works will focus on implementing similar analyses for other color sampling mechanisms such as 4:2:2 or 4:2:0, and to design deep learning schemes dedicated to color JPEG images.

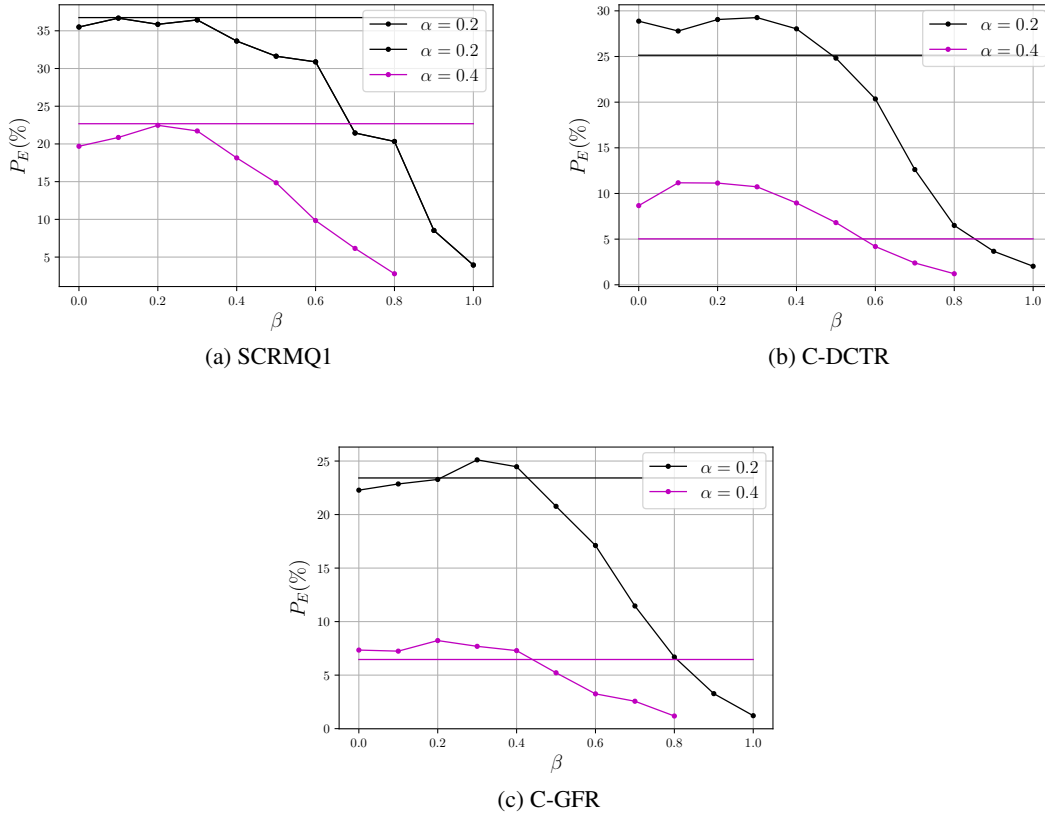


Fig. 3: UERD: Comparison w.r.t. β for different feature sets, JPEG QF = 75. Horizontal lines are results for the CONC strategy.

6 Acknowledgments

The authors would like to thank Rémi Duprès, who designed the pipeline used to perform this set of extensive tests. This work was partially supported by the French ANR DEFALS program (ANR-16-DEFA-0003).

References

1. Hasan Abdulrahman, Marc Chaumont, Philippe Montesinos, and Baptiste Magnier, “Color image steganalysis based on steerable gaussian filters bank,” in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*. ACM, 2016, pp. 109–114.
2. Miroslav Goljan and Jessica Fridrich, “Cfa-aware features for steganalysis of color images,” in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2015, pp. 94090V–94090V.
3. Miroslav Goljan, Jessica Fridrich, Rémi Cogranne, et al., “Rich model for steganalysis of color images,” in *Parallel Computing Technologies (PARCOMPTECH), 2015 National Conference on*. IEEE, 2015, pp. 185–190.

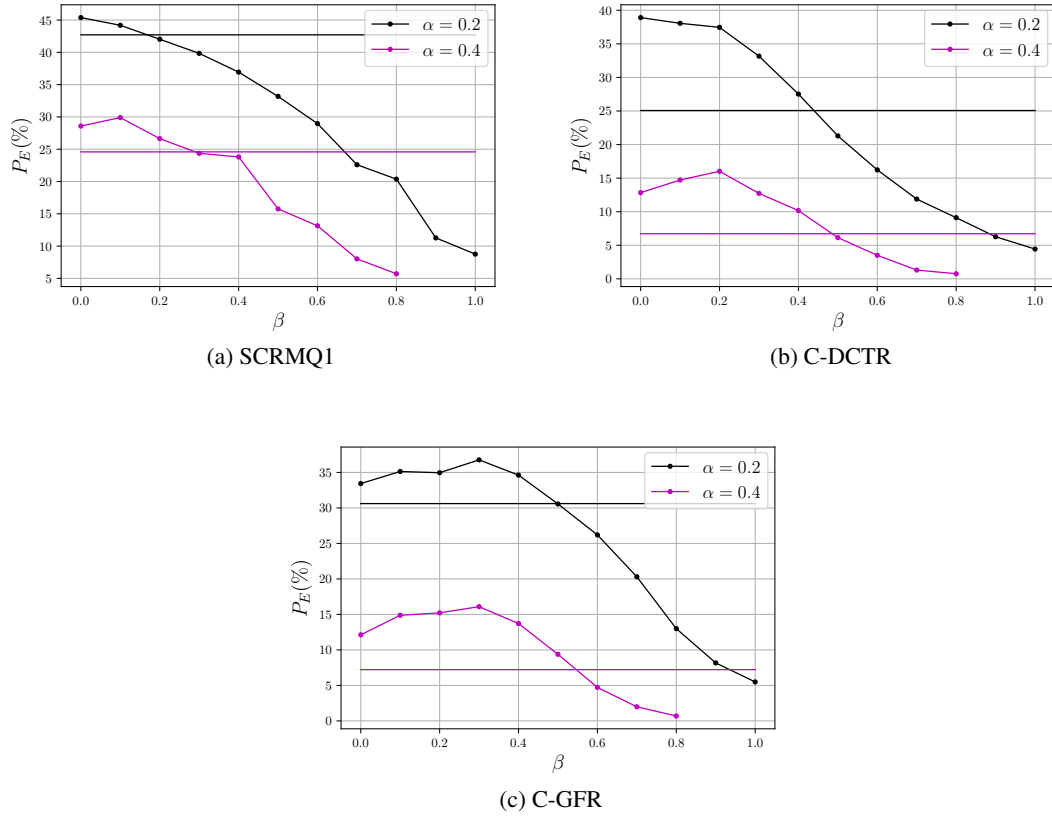


Fig. 4: J-UNIWARD: Comparison w.r.t. β for different feature sets, JPEG QF = 95. Horizontal lines are results for the CONC strategy.

4. Bin Li, Ming Wang, Xiaolong Li, Shunquan Tan, and Jiwu Huang, “A strategy of clustering modification directions in spatial image steganography,” *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 9, pp. 1905–1917, 2015.
5. Andreas Westfeld, “High capacity despite better steganalysis: F5- a steganographic algorithm,” in *Fourth Information Hiding Workshop*, 2001, pp. 301–315.
6. Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark, “Universal distortion function for steganography in an arbitrary domain,” *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.
7. Andrew D Ker, Patrick Bas, Rainer Böhme, Rémi Cogranne, Scott Craver, Tomáš Filler, Jessica Fridrich, and Tomáš Pevný, “Moving steganography and steganalysis from the laboratory into the real world,” in *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 45–58.
8. Jessica Fridrich and Jan Kodovsky, “Rich models for steganalysis of digital images,” *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 868–882, 2012.
9. Jishen Zeng, Shunquan Tan, Guangqing Liu, Bin Li, and Jiwu Huang, “Wisernet: Wider separate-then-reunion network for steganalysis of color images,” *arXiv preprint arXiv:1803.04805*, 2018.

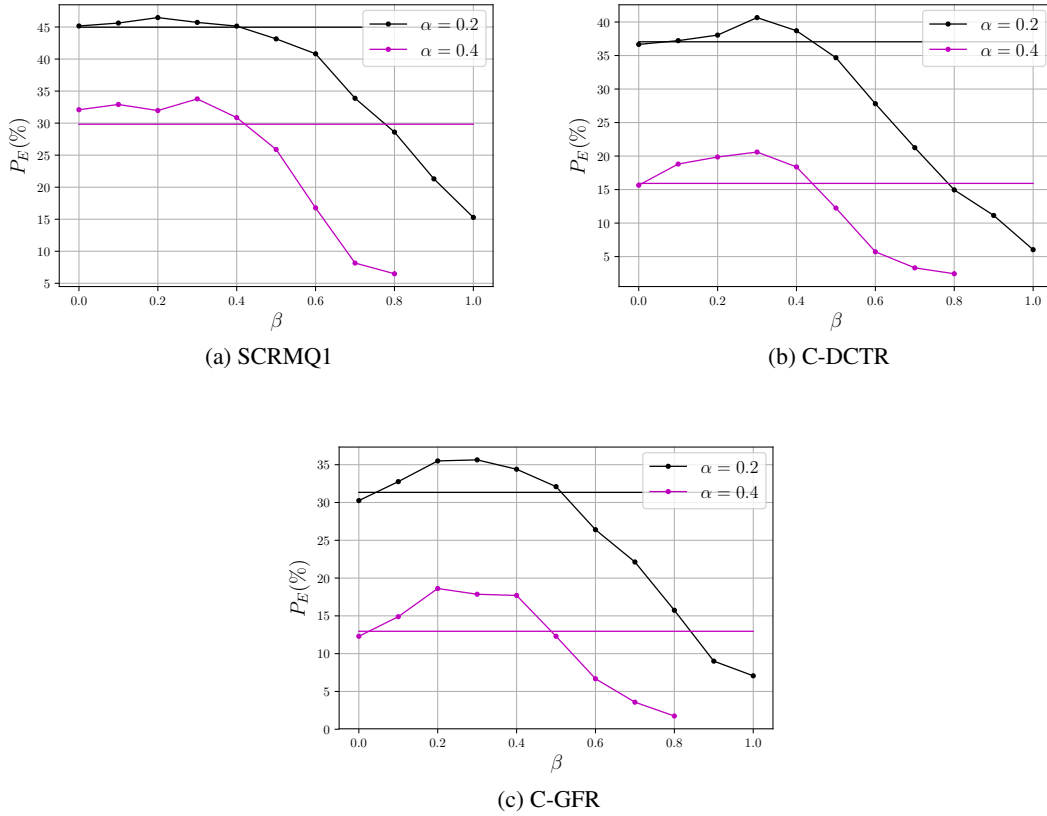


Fig. 5: UERD: Comparison w.r.t. β for different feature sets, JPEG QF = 95. Horizontal lines are results for the CONC strategy.

10. Vojtěch Holub and Jessica 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.
11. Xiaofeng Song, Fenlin Liu, Chunfang Yang, Xiangyang Luo, and Yi Zhang, “Steganalysis of adaptive jpeg steganography using 2d gabor filters,” in *Proceedings of the 3rd ACM workshop on information hiding and multimedia security*. ACM, 2015, pp. 15–23.
12. Linjie Guo, Jiangqun Ni, Wenkang Su, Chengpei Tang, and Yun-Qing 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.
13. A. Ker, “Batch steganography and pooled steganalysis,” in *Information Hiding*. Springer, 2007, pp. 265–281.
14. P. Bas, T. Pevný, and T. Filler, “Bossbase,” <http://exile.felk.cvut.cz/boss>, May 2011.
15. Jan Kodovsky, Jessica Fridrich, and Vojtech Holub, “Ensemble classifiers for steganalysis of digital media,” *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, 2012.
16. Andrew D Ker, Tomáš Pevný, Jan Kodovský, and Jessica Fridrich, “The square root law of steganographic capacity,” in *Proceedings of the 10th ACM workshop on Multimedia and security*. ACM, 2008, pp. 107–116.

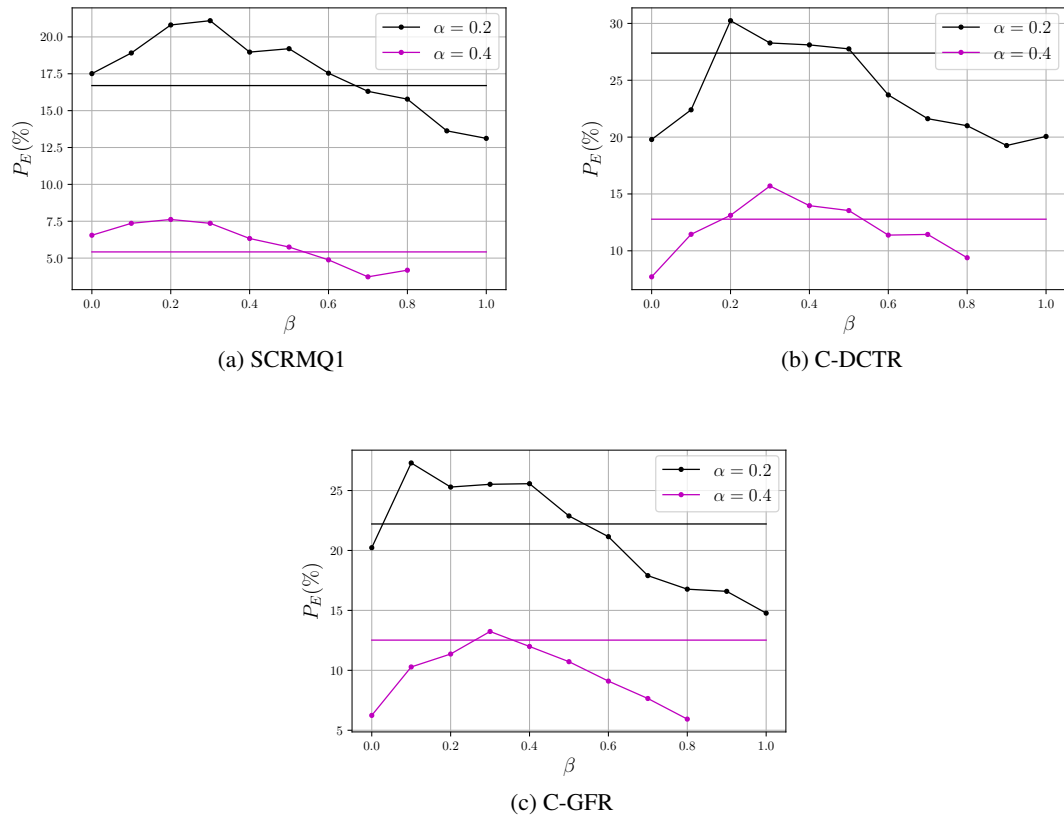


Fig. 6: J-UNIWARD: Comparison w.r.t. β for different feature sets, JPEG QF = 100. Horizontal lines are results for the CONC strategy.

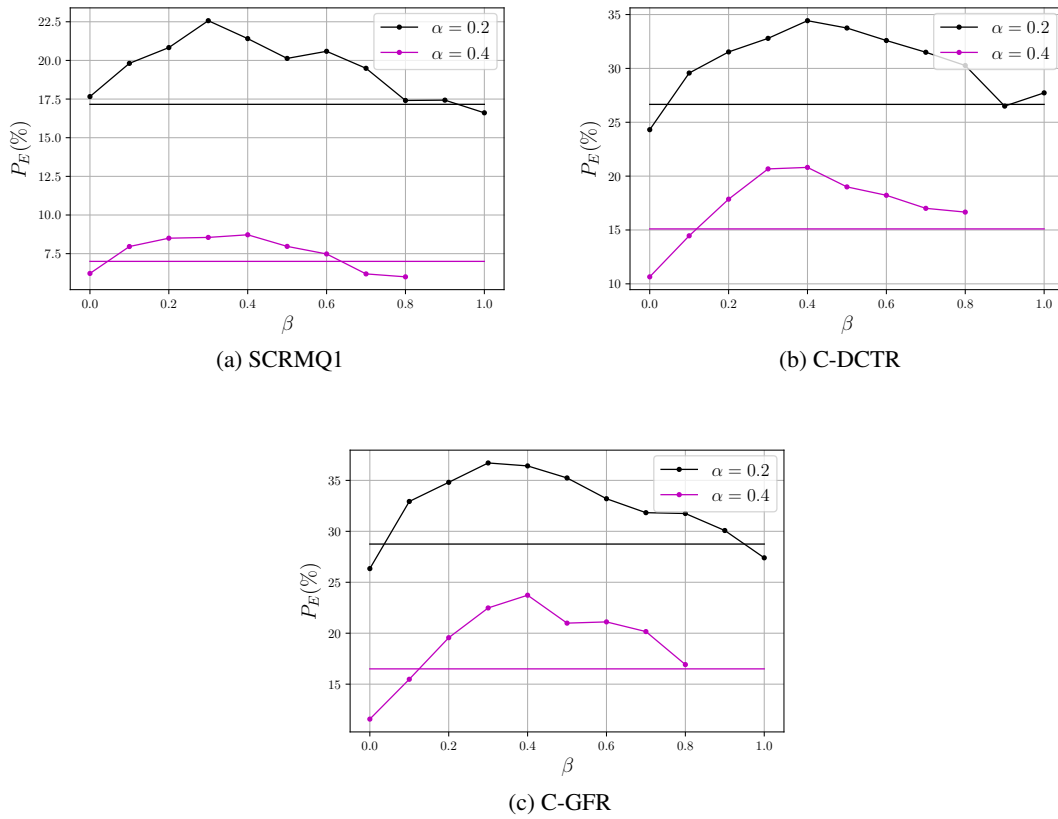


Fig. 7: UERD: Comparison w.r.t. β for different feature sets, JPEG QF = 100. Horizontal lines are results for the CONC strategy.