Supplementary Materials for: Evaluation of Random Field Models in Multi-modal Unsupervised Tampering Localization

Paweł Korus and Jiwu Huang

This document extends upon selected aspects of the work described in manuscript *Evaluation of Random Field Models in Multi-modal Unsupervised Tampering Localization* presented at IEEE International Workshop on Information Forensics and Security, 2016, Abu Dhabi.

The document is available online from: http://kt.agh.edu.pl/~korus/publications/2016-wifs

I. Implementation Details

Localization algorithms: Our work used two popular forensic detectors: (a) state-of-the-art CFA detector [1] with publicly available implementation[1]; (b) standard PRNU detector with a correlation predictor [2]. The latter has been modified to yield tampering probability maps $\mathbf{c}$ according to the Bayesian formulation:

$$c_i = \mathcal{B}(q_i|\sigma_0,\sigma_1,\hat{q}_i) = \frac{P_{\mathcal{N}(0,\sigma_0)}(q_i)}{P_{\mathcal{N}(\hat{q}_i,\sigma_1)}(q_i) + P_{\mathcal{N}(0,\sigma_0)}(q_i)} \quad (1a)$$

$$= \left(1 + e^{-\log(\sigma_1/\sigma_0) - \frac{(q_i-\hat{q}_i)^2}{2\sigma_1^2}) + \frac{q_i^2}{2\sigma_0^2}}\right)^{-1} \quad (1b)$$

where q_i is the measured correlation for a square window centered around pixel i, $\hat{q}_i$ is the expected correlation value estimated by a standard predictor, and σ_0, σ_1 are the variances of Gaussian models assumed for hypothesis H_0 (signature absent) and H_1 (signature present). More information about our implementation can be found in [3]. Source code is available online[2].

Both detectors were configured to yield tampering probability maps of the same size. According to the recommendations from the original paper [1], we used block aggregation to obtain localization resolution of 8×8 px. We set analysis window stride to obtain analogous resolution for the PRNU detector. As a result, for 1920×1080 px images, the detectors yielded tampering probability maps of size 240×135 px. In the final stage, the obtained decision maps were resized to the original image size for post-processing (if needed) and comparison with pixel-level ground truth tampering maps.

Empirical Fusion: The *empirical* fusion method discussed in our paper can be seen as a variation of the *behavioral knowledge space* combination rule [4]. The obtained rule is visualized in Fig. 3 in the manuscript and numerical values for the rule are reported as Tab. I. Our implementation used local regression smoothing (the *loess* surface fit in Matlab) to generalize the rule to arbitrary real-valued inputs.

By including this method, our goal was to illustrate general qualitative differences between commonly used naive combination rules and empirical behavior supported by real data. The presented empirical rule has enforced symmetry in order to show general trends for possibly arbitrary candidate detectors. We note that quantitatively slightly better results may be obtained if the symmetry constraint is dropped (although in general this will depend on the detectors in the ensemble at hand).

Random Field Models: We used existing solvers to find the optimal tampering maps according to the *grid* and *dense* CRF models. We used a graph cuts-based solver [5] from the UGM toolbox [6] to quickly find the optimal tampering map for the *grid* CRF. For the *dense* CRF we used a recently proposed efficient solver[3] based on iterative mean-field approximations [7]. For the sake of research reproducibility, our implementations of the discussed fusion methods are available as supplementary materials online.

TABLE I
Empirical Fusion Table

0.0027	0.0060	0.0075	0.0128	0.0186	0.0405	0.1067	0.4678
0.0060	0.0150	0.0163	0.0186	0.0315	0.0600	0.1927	0.6840
0.0075	0.0163	0.0244	0.0314	0.0582	0.1267	0.3214	0.7512
0.0128	0.0186	0.0314	0.0459	0.0834	0.1951	0.4298	0.8789
0.0186	0.0315	0.0582	0.0834	0.1467	0.3781	0.6740	0.9475
0.0405	0.0600	0.1267	0.1951	0.3781	0.6384	0.9012	0.9875
0.1067	0.1927	0.3214	0.4298	0.6740	0.9012	0.9733	0.9973
0.4678	0.6840	0.7512	0.8789	0.9475	0.9875	0.9973	0.9985

TABLE II
Interpretation of the parameters of the CRFs

Parameter	Meaning
β_0	default neighborhood interaction strength
β_1	content-dependent interaction strength
θ_0	fall-off strength for the spatial distance in default neighborhood interactions (dense CRF)
θ_1	fall-off strength for the spatial distance in content-dependent neighborhood interactions (dense CRF)
θ_2	fall-off strength for the color distance in content-dependent neighborhood interactions

II. Parameter Selection and Threshold Sensitivity

The conditional random fields used in our study are controlled by only a few parameters. While these parameters may be learned from the data (e.g., by loss driven grid / random search), reasonable values may be chosen by trial and error. The parameters have clearly defined impact on the behavior of the localization algorithm (Tab. II) and may be manually fine-tuned for individual cases.

Compared to other fusion methods, it can be observed that the CRF fusion delivered its best performance (at least F_1 score-wise) for a relatively narrow range of threshold values. For standard supervised use cases it may not be a critical problem, since the optimal threshold is close to 0.5, and forensic analysts are likely to adjust the threshold anyway. In general however, and particularly in unsupervised applications, it may be preferable to increase the range of high-performing threshold settings (even at the cost of peak performance). In the described experiments, we aimed to

[1] http://lesc.det.unifi.it/en/node/187

[2] https://github.com/pkorus/multiscale-prnu

[3] http://www.philkr.net/home/densecrf

maximize the average F_1 score, regardless of the threshold. We have performed preliminary experiments with alternative objective functions where all thresholds are taken into account but with different weights (high weights around $\tau = 0.5$, low otherwise). This approach proved to be unsuccessful leading to deteriorated peak performance and nearly unaffected range of high-performing thresholds.

We have also experimented with alternative formulations of the random field. In the current study, we constructed the unary potentials based on *quasi-thresholds*, leading to the following energy function:

$$E_Q(\mathbf{t}) = \frac{1}{|D|} \sum_{d \in \mathcal{D}} \sum_{i=1}^{N} \psi_\tau(c_i^{(d)}|t_i) + \sum_{i=1}^{N} \sum_{j \in \Xi_i} \phi_p(t_i, t_j) \quad (2)$$

where ψ_τ, and ϕ_p denote the unary and pairwise potentials, respectively, and Ξ_i denotes the neighborhood of pixel i. We use the approach from [8] to construct the unary potentials:

$$\psi_\tau(c|t) = -\log \max\left(\Psi_{\text{min}}, \Psi_\tau(c|t)\right), \quad (3)$$

with $\Psi_{\text{min}} \in [0, 1]$ and:

$$\Psi_\tau(c|t) = \begin{cases} 1 - \frac{c}{2\tau} & \text{for } t = 0, \\ 1 + \frac{c}{2(1-\tau)} - \frac{1}{2(1-\tau)} & \text{for } t = 1, \end{cases} \quad (4)$$

where $\tau \in (0, 1)$ is a quasi-threshold that equalizes potentials for both decisions, i.e., $\psi_\tau(\tau, 0) = \psi_\tau(\tau, 1)$. Hence, in this case the quasi-threshold serves as a control variable that adjusts the bias of the detector. If neighborhood interaction strengths are set to 0, this construction is equivalent to a standard decision threshold.

Alternatively, one may construct a CRF model with standard Ising potentials [9], where the energy function assumes the following form:

$$E_I(\mathbf{t}) = \frac{1}{|D|} \sum_{d \in \mathcal{D}} \sum_{i=1}^{N} \psi_{0.5}(c_i^{(d)}|t_i) \quad (5a)$$

$$+ \frac{\alpha}{2} \sum_{i=1}^{N} \mathcal{I}(t_i = 1) \quad (5b)$$

$$+ \sum_{i=1}^{N} \sum_{j \in \Xi_i} \phi_p(t_i, t_j) \quad (5c)$$

where $\mathcal{I}(t_i = 1)$ is an indicator function:

$$\mathcal{I}(C) = \begin{cases} 1 & \text{if } C \text{ is true}, \\ -1 & \text{otherwise}, \end{cases} \quad (6)$$

and $\alpha = \log \frac{p}{1-p}$. The decision bias is hence controlled by the parameter $p \in [0, 1]$ which denotes the prior probability of the pixels being tampered. In this case, the quasi-threshold is fixed at 0.5 and adjusting p corresponds to penalizing imbalance between the labels.

A comparison of the quasi-threshold and Ising variants of the CRF is shown in Fig. 1. In both cases, we used the same neighborhood interaction strengths ($\beta_0 = 0.25, \beta_1 = 2$). As expected, there are no differences in peak performance and both methods return equivalent results for $\tau = p = 0.5$ since:

$$E_Q(\mathbf{t}|\tau = 0.5) = E_I(\mathbf{t}|p = 0.5) \quad (7)$$

It can be observed that the standard Ising model maintains greater F_1 scores for slightly wider range of the control variable (τ or p, respectively) - especially for values above 0.5. However, the receiver operation characteristic favors the quasi-threshold model.

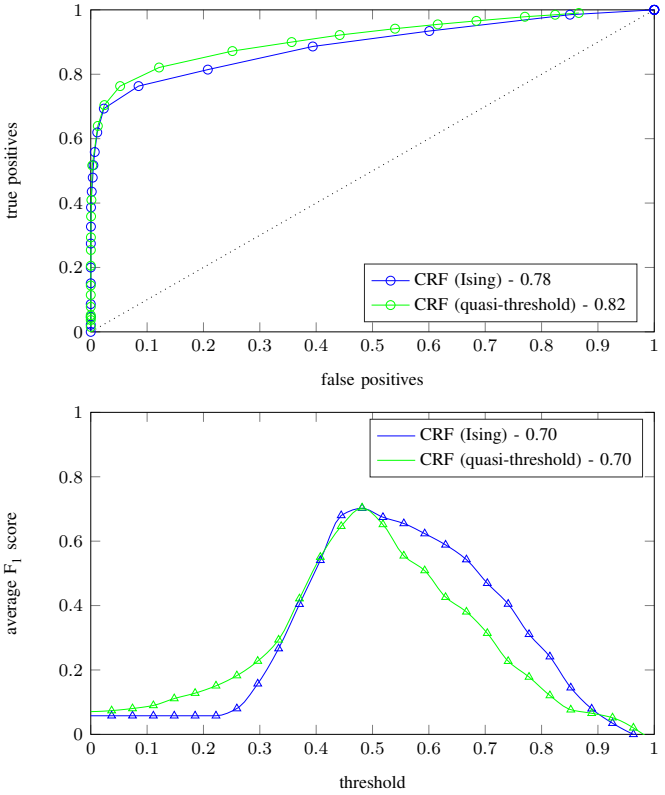

Fig. 1. Comparison of localization performance for CRF formulation with quasi-thresholds (green) and with standard Ising potentials (blue); receiver operation characteristic (top); F_1 scores (bottom).

III. CRF for Single-Modality Localization

Conditional random fields can also be used with a single forensic detector. Adoption of neighborhood interactions eliminates the need for ad-hoc heuristic post-processing and can further improve shape representation when the interaction strength is determined in a content-adaptive way. In order to illustrate this, we have repeated our evaluation of the grid CRF for a single detector scenario, i.e., for $\mathcal{D} = \{\text{cfa}\}$ and $\mathcal{D} = \{\text{prnu}\}$. Parameters of the CRF were left unchanged ($\beta_0 = 0.25, \beta_1 = 2$). The baseline post-processing heuristics involved morphological opening (CFA detector) and small connected component removal combined with morphological dilatation (PRNU detector).

The results are reported in Fig. 2. Successive plots correspond to receiver operation characteristics (1st column), dependencies of F_1 scores on the decision threshold τ (2nd column), and scatter plots of per-image peak F_1 scores (3rd

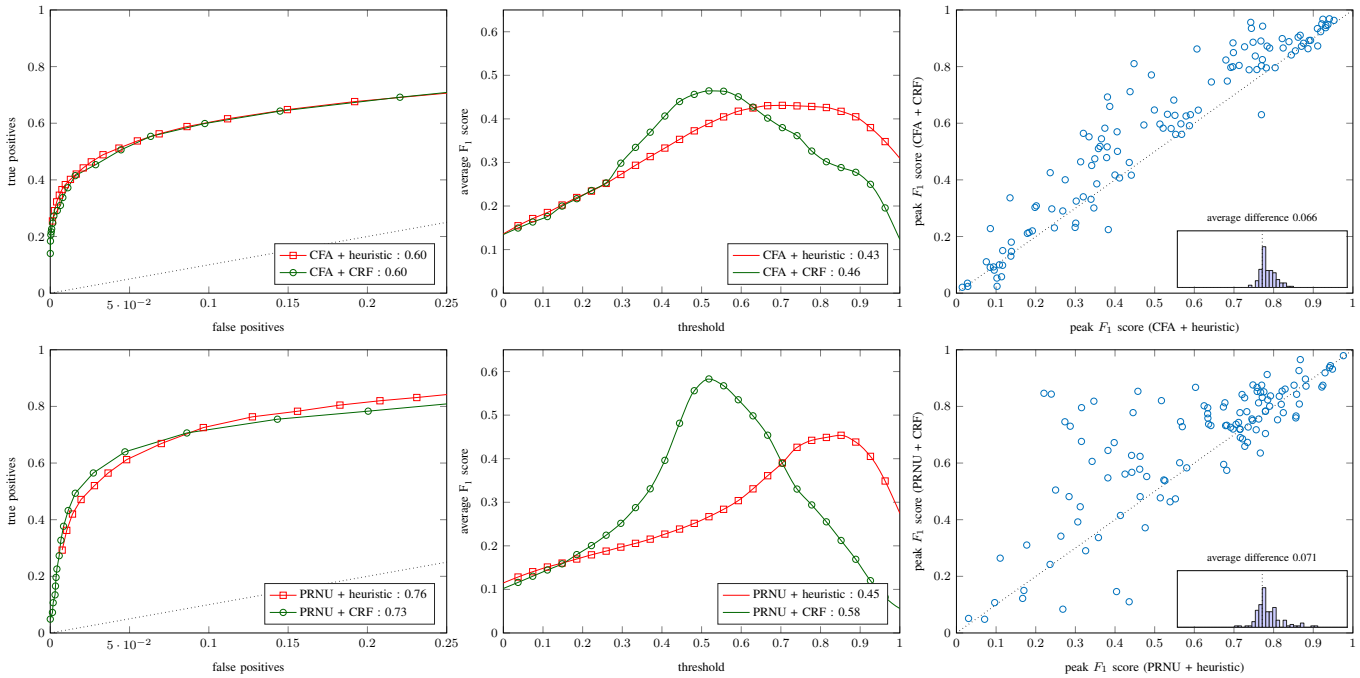

Fig. 2. Comparison of tampering localization results of individual detectors for a conditional random field model against standard heuristic post-processing: CFA with morphological opening heuristic (top); PRNU with small connected component removal and morphological dilatation (bottom).

column). For the PRNU detector (bottom row), we observed a minor improvement of the ROC curve for low false positive rates fp which starts to deteriorate around $fp \gtrapprox 0.09$. The F_1 score reveals a significant performance improvement. Both the average and peak scores show clear benefits of the CRF model. For the CFA detector, the average scores reveal practically no improvement. We observed a minor increase of the average F_1 score and a nearly identical ROC curve. Interestingly, individual peak F_1 scores show considerable advantages of the CRF model when the decision threshold does not need to be identical for all images. This indicates that when manual fine-tuning of the threshold / parameters is feasible, the CRF model can bring even more benefits.

Finally, we note that it may be possible to improve ROC performance by redefining the parameter selection objective. In case of manual parameter selection, the choice was guided by our subjective perception of the localization utility. In cross-validation (discussed in the manuscript), the choice was driven by the best average F_1 score. In principle, the objective can be easily changed to ROC-based metrics, like the AUC. However, at this moment it is not clear which performance metric is better. As demonstrated in the paper, they are both inadequate to assess the performance of precise tampering localization.

IV. Limitations & Future Work

Our current evaluation covered a case of bi-modal analysis with two distinct detectors. However, our framework supports an arbitrary number of candidate maps, and good performance can be expected for many detectors with similar output. Our current formulation requires the algorithms to return real-valued tampering probability maps which are a convenient common denominator. Such output is easily understandable by humans and is therefore generated by many schemes, including state-of-the-art double JPEG compression detectors, e.g., [10]. However, further work would be required to make the algorithm more general. Some detectors may operate on even larger blocks (e.g., resampling detectors often need windows of size at least 256×256 px) or in a completely different domain (e.g., image segments or super-pixels). In such conditions, it may be worth to consider constructing a hierarchical CRF [11].

The presented framework could also be extended to support detectors that do not yield probabilistic output. Examples of such algorithms include JPEG ghosts [12] or JPEG quantization table estimation techniques [13]. The unary potentials would need to be redesigned to map the algorithm output to the energies of individual labels. Further work would also be required to introduce prior knowledge about detector compatibility, similarly to full-frame decision frameworks based on DSTE [14] or fuzzy logic [15]. Analogous functionality may be obtained by redefining unary potentials to yield uncertain states (similar energies for both labels) if conflicting traces are found. The neighborhood interactions would then be able to propagate confident scores from surrounding regions. Alternatively, an additional label may be introduced to indicate such conflicting areas. The issue requires a more thorough, dedicated study.

V. Grid vs. Dense CRF

Due to limited capabilities of existing evaluation metrics, our study did not reveal performance benefits of the more complex dense CRF. While this approach can delineate objects

more accurately than the simpler grid CRF, the improvement is not reflected in either classification accuracy or F_1 scores. Without clear measures of the actual localization map utility, we can only conclude that the benefits are mostly aesthetic. Given that this approach requires to set a larger number of parameters, the simpler grid CRF may be a more practical choice. While even larger number of parameters can be efficiently learned with existing algorithms [16], the lack of suitable objective functions makes it a challenging problem.

REFERENCES

[1] P. Ferrara, T. Bianchi, A. Rosa, and A. Piva, "Image forgery localization via fine-grained analysis of cfa artifacts.," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1566–1577, 2012.

[2] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 74–90, 2008.

[3] P. Korus and J. Huang, "Multi-scale analysis strategies in PRNU-based tampering localization," *IEEE Trans. Inf. Forensics Security*, 2017 (To appear).

[4] Ludmila I Kuncheva, *Combining pattern classifiers: methods and algorithms*, John Wiley & Sons, 2004.

[5] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[6] M. Schmidt, "UGM: A matlab toolbox for probabilistic undirected graphical models," http://www.cs.ubc.ca/~schmidtm/Software/UGM.html, 2011 version.

[7] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Proc. of NIPS*, 2011.

[8] P. Korus and J. Huang, "Multi-scale fusion for improved localization of malicious tampering in digital images," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1312–1326, 2016.

[9] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A Bayesian-MRF approach for PRNU-based image forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 554–567, 2014.

[10] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of jpeg artifacts," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 1003–1017, 2012.

[11] C. Russell, P. Kohli, P. HS Torr, et al., "Associative hierarchical CRFs for object class image segmentation," in *IEEE Int. Conf. Computer Vision*, 2009.

[12] H. Farid, "Exposing digital forgeries from jpeg ghosts," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 154–160, 2009.

[13] B. Li, T.-T. Ng, X. Li, S. Tan, and J. Huang, "Statistical model of JPEG noises and its application in quantization step estimation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1471–1484, 2015.

[14] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A Framework for Decision Fusion in Image Forensics Based on Dempster-Shafer Theory of Evidence," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 4, pp. 593–607, 2013.

[15] M. Barni and A. Costanzo, "Dealing with uncertainty in image forensics: a fuzzy approach," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2012, pp. 1753–1756.

[16] P. Krähenbühl and V. Koltun, "Parameter learning and convergent inference for dense random fields," in *Proc. of ICML*, 2013.