Bachelor Thesis

Comparison of Hamming- and Variation of Information-Loss based structured learning on the Multicut Problem

Jan Lammel

19. Februar 2016

Table of Contents

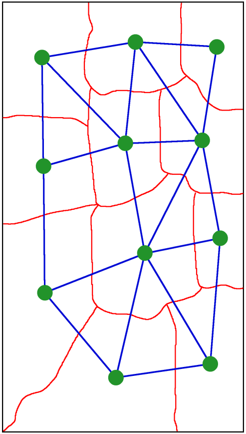
Segmentation

Motivation Variation of Information

- Hamming Loss strongly dependend on exact path of segmentation
- But: Path of segmentation often not unique

- Idea VOI: Consider labels of segmentation and penanlize area-dependend

Region Adjacency Graph (RAG)



- Image partitioned into Superpixel (SP) via SLIC [2]
- Each Superpixel $\hat{=}$ Node in RAG
- Nodes of adjacent SP are linked by an Edge

Multicut Problem (MP)

$$
\begin{aligned}
\min_{y} \quad & \sum_{y_e \in E} \langle w, \beta_e \rangle \cdot y_e \\
\text{s.t.} \quad & y - \sum_{y_i \in P(y)} y_i \leq 0 \qquad \forall\, y \in E
\end{aligned}
$$

- w: Weights to be learned
- β_e: Features of edge e
- y_e: Activity of edge e
- Constraint to enforce consistency

Stochastic Gradient with RF Feature

- Varying configurations:
  - Domain Feature Space
  - Constraint on RF Feature
  - Subgradient Descent with/without RF Feature
- Results:
  - Decrease of VOI Loss leads to decline of Hamming Loss in Trainingsset
  - Rate of decrease sensible to configuration, besides strong fluctuations due to stochastic process
  - Loss decrease on Trainingset $\propto$ Loss increase on Testset $\Rightarrow$ Overfit of training data

Stochastic Gradient without RF Feature

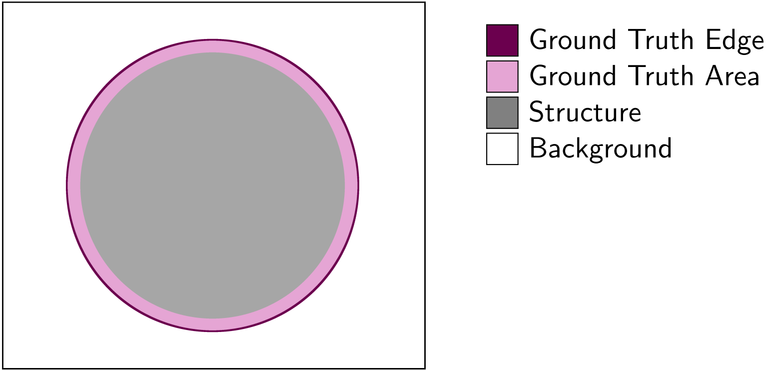
- Varying configurations:
  - Domain Feature Space
  - Constraint on N^4 Feature
- Results:
  - VOI Loss decrease on Trainingset of approximately 4%
  - Change of Loss on Testset within 1σ range of error

Cross Validation Measurement 10

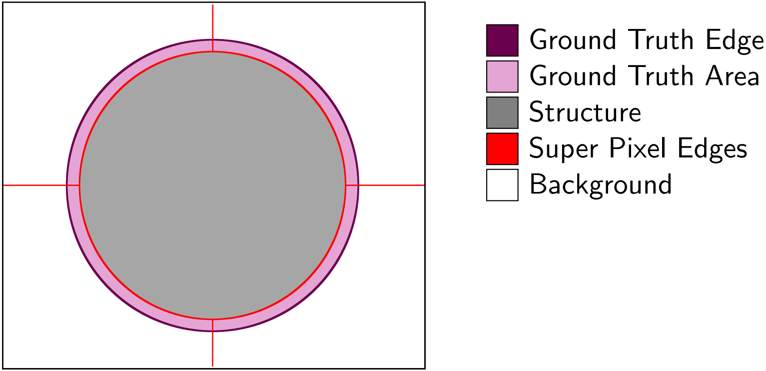
- Cross validation to minimize measurement errors
- Results on Testset:
  - $\mathcal{L}_H$: $\frac{StochGrad}{SubGrad} = 0.989 \pm 0.005$
  - $\mathcal{L}_{VOI}$: $\frac{StochGrad}{SubGrad} = 1.0025 \pm 0.0084$

  $\rightarrow$ No significant change

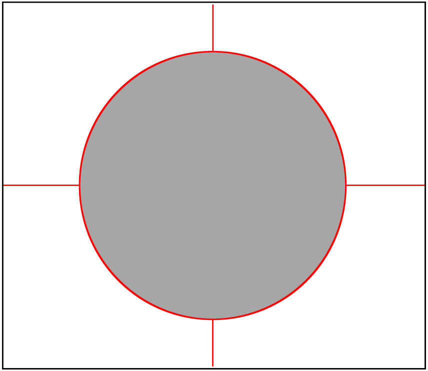
Explanation by SLIC I



Explanation by SLIC I

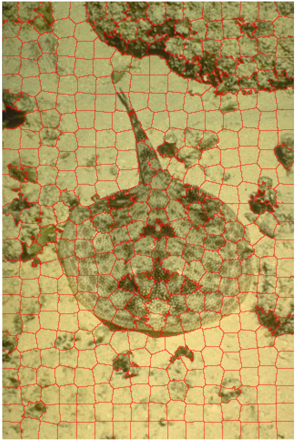


Explanation by SLIC I



Structure
Super Pixel- &
Ground Truth Edges

Explanation by SLIC I



Conclusion

- Stochastic Gradient with RF Feature leads to Overfit of training data
- No significant change without RF Feature
  - Bad Ground Truth compensated by SLIC
  - SLIC provides just important edges
    $\rightarrow$ Difficulty of exact segmentation path is gone
  - Examination on Pixel-level would be interesting

[1] Sebastian Nowozin, Christoph H. Lampert, *Structured Learning and Prediction in Computer Vision*, 2011.

[2] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, *SLIC Superpixels Compared to State-of-the-art Superpixel Methods*, 2011

[3] D. Martin, C. Fowlkes, D. Tal and J. Malik, *A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics*, Proc. 8th Int'l Conf. Computer Vision, July 2001

[4] Yaroslav Ganin and Victor Lempitsky, *N^4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms*, Skolkovo Institute of Science and Technology, 2014

[5] Piotr Dollár and C. Lawrence Zitnick, *Structured Forests for Fast Edge Detection*, ICCV 2013

[6] Ullrich Köthe, *Vision with Generic Algorithms*, Image Processing and Analysis Library, Version 1.10.0

[7] Thorsten Beier, *Inferno*, https://github.com/DerThorsten/inferno

[8] P. Arbeláez, M. Maire, C. Fowlkes and J. Malik, *Contour Detection and Hierarchical Image Segmentation*, IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011

[9] Saining Xie and Zhuowen Tu, *Holistically-Nested Edge Detection*, CoRR, abs/1504.06375, 2015

[10] Bjoern Andres, Joerg H. Kappes, Thorsten Beier, Ullrich Köthe and Fred A. Hamprecht, *Probabilistic Image Segmentation with Closedness Constraints* ICCV, Proceedings, (2011), 2611 - 2618

[11] Jörg Hendrik Kappes, Markus Speth, Björn Andres, Gerhard Reinelt, Christoph Schnörr, *Globally Optimal Image Partitioning by Multicuts*, EMMCVPR, Springer, 2011

[12] Jörg Hendrik Kappes, Markus Speth, Gerhard Reinelt, Christoph Schnörr, *Higher-order Segmentation via Multicuts*, 2013

[13] Kappes, J.; Swoboda, P.; Savchynskyy, B.; Hazan, T. and Schnörr, C., *Probabilistic Correlation Clustering and Image Partitioning Using Perturbed Multicuts*, In Proc. SSVM, Springer, LNCS , 2015

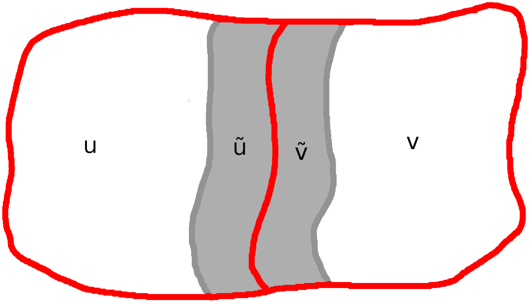
[14] Margret Keuper, Evgeny Levinkov, Nicolas Bonneel, Guillaume Lavoué, Thomas Brox, Bjoern Andres, *Efficient Decomposition of Image and Mesh Graphs by Lifted Multicuts*, eprint arXiv:1505.06973, 05/2015

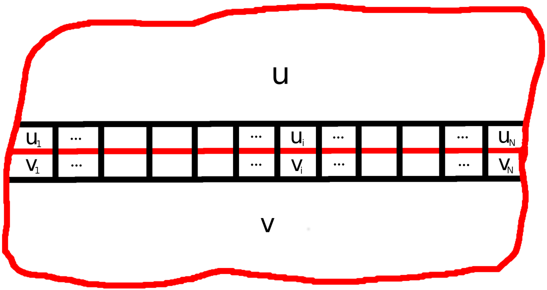
[15] Sebastian Nowozin and Christoph H. Lampert, *Global Connectivity Potentials for Random Field Models*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)

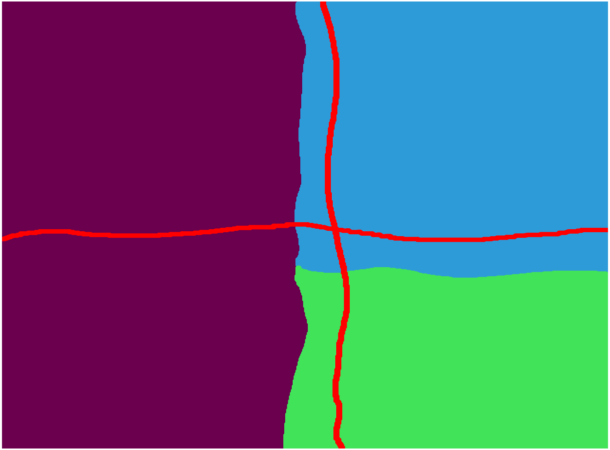
[16] S. Beucher and F. Meyer, *Mathematical Morphology in Image Processing*, Marcel Dekker, 1992, ch. 12

[17] Jordi Pont-Tuset and Pablo Arbelaez and Jonathan T. Barron and Ferran Marqués and Jitendra Malik, *Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation*, CoRR, abs/1503.00848, 2015

Questions


u
ũ
ṽ
v


u
u1
...
...
ui
...
...
uN
v1
...
...
vi
...
...
vN
v



Algorithm 1 Get Gradient Descent Direction

procedure GetGradientDescentDirection(#Perturbs, σ, w)
    σ: Noise standard deviation
    w: Current Weight Wector

    $\Delta x = 0$
    for $n = 1...\#\text{Perturbs}$ **do**
        Generate Noise $\in \mathcal{N}(0, \sigma^2)$ und add to w
        Calculate Loss on current Training Sample
        $\Delta x = \Delta x + \text{Noise} * \text{Loss}$
    $\Delta x = -\Delta x / \#\text{Perturbs}$
    return Δx

Algorithm 2 Line Search and update Weights

procedure LINESEARCHANDTAKESTEP($\eta, It, w, \Delta x, \Delta w_{prev}$)

  η: Stepwidth

  It: Current Iteration

  w: Current weight vector

  Δx: Gradient Descent Direction

  Δw_{prev}: Step of previous iteration

  $\eta_{eff} = \eta / It$

  for $n = \{0.1, 0.5, 1.0, 5.0, 10.0\}$ **do**

    Varied Weight Vector $w_{var} = w + \eta_{eff} \cdot \Delta x \cdot n + m \cdot w_{prev}$

    Calculate mean Loss $\mathcal{L}$ on entire Training Set

    from w_{var}

    if $\mathcal{L} < \mathcal{L}_{best}$ **then**

      $\mathcal{L}_{best} = \mathcal{L}$

      Save $w_{best} = w_{var}$

      Break

    Memorize $\mathcal{L}$ and associated varied weight vector

  $w = w_{var}$, where regarding Loss is minimal

  return w

$$\mathcal{L}_i(y_i, y_i^*) = \begin{cases} \mathbb{I}[y_i \neq y_i^*] \cdot \alpha_{over} & \text{if} \quad y_i^* = 0 \\ \mathbb{I}[y_i \neq y_i^*] \cdot \alpha_{under} & \text{if} \quad y_i^* = 1 \end{cases} \qquad \forall y_i \in E$$

$$\mathcal{L}_H(y, y^*) = \sum_{y_i \in E} \mathcal{L}_i(y_i, y_i^*)$$

$$\mathcal{L}_{\mathrm{VOI}}(y, y^*) = H_y + H_{y^*} - 2 \cdot I(y, y^*)$$

$$H_y = \mathbb{E}[\hat{I}(y)] = -\sum_{l \in y} p(l) \cdot \log_e(p(l))$$

$$I(y, y^*) = \sum_{l_1 \in y} \sum_{l_2 \in y^*} p(l_1, l_2) \cdot \log_e \left(\frac{p(l_1, l_2)}{p(l_1)p(l_2)} \right)$$