# Image Captioning to Assist the Visually Impaired

PIERRE DOGNIN, IGOR MELNYK, YOUSSEF MROUEH, INKIT PADHI, MATTIA RIGOTTI, JERRET ROSS, YAIR SCHIFF*

*CVPR VizWiz Grand Challenge Workshop*
*June 14, 2020*

*All authors contributed equally

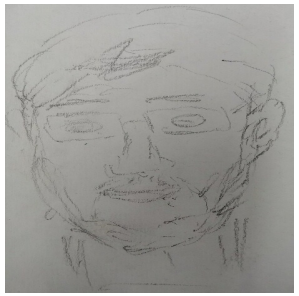
**Pierre Dognin**

**Igor Melnyk**

**Youssef Mroueh**

**Inkit Padhi**

**Mattia Rigotti**

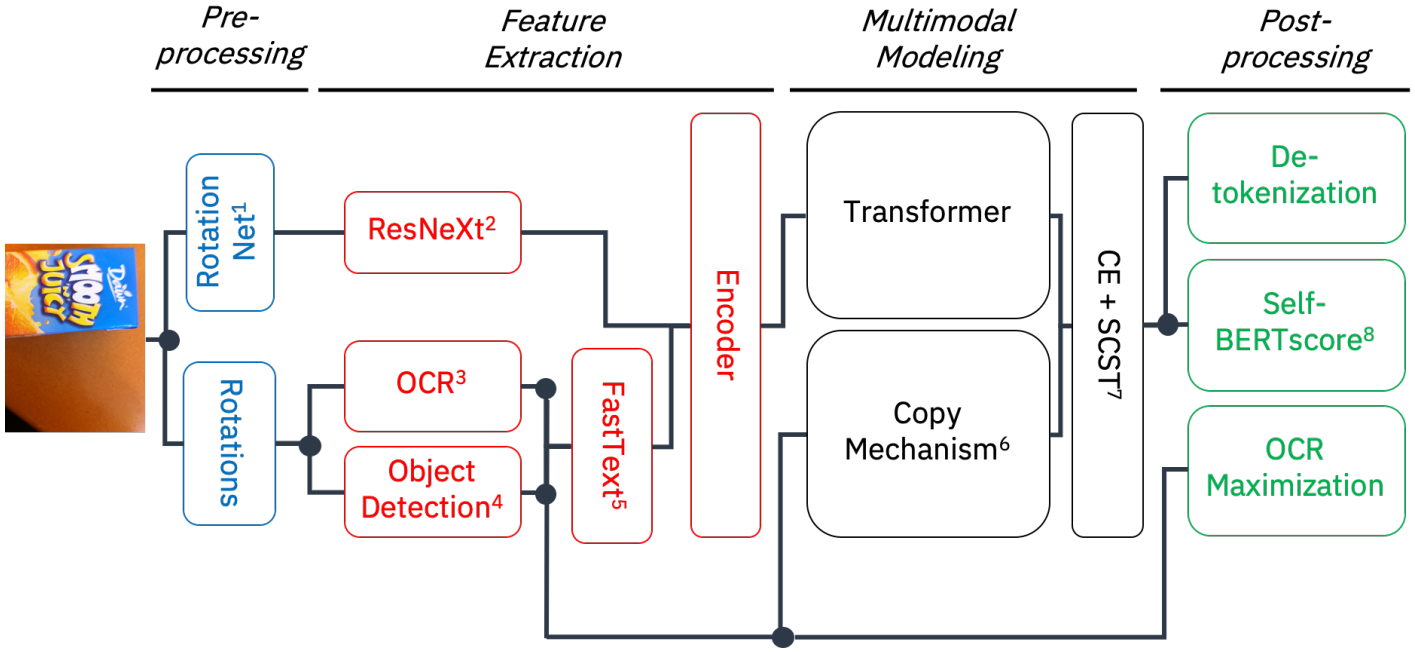


**Jerret Ross**

**Yair Schiff**

**Richard Young**

# MODEL:




# MAIN IDEAS:

Open Source

Ensemble

Pre/Post-process

Icon credits: Open source: https://icons8.com/icon/65836/open-source Org Chart: https://webstockreview.net/explore/organization-clipart-org-chart/ Paintbrush: https://www.clipartkey.com/view/bwhhbo_transparent-paintbrush-clipart-black-and-white-paint-brush/

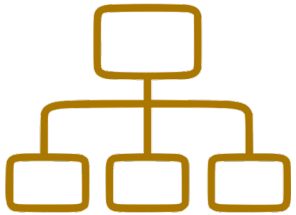


1. Spyros Gidaris, Praveer Singh, and Nikos Komodakis. “Unsupervised Representation Learning by Predicting Image Rotations”. In: CoRRabs/1803.07728(2018). arXiv: 1803.07728. URL: http://arxiv.org/abs/1803.07728
2. Saining Xie et al. “Aggregated Residual Transformations for Deep Neural Networks”. In: arXiv preprint arXiv: 1611.05431(2016)
3. Jeonghun Baek et al. “What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis”. In: International Conference on Computer Vision (ICCV). to appear. 2019. published.
   Youngmin Baek et al. “Character Region Awareness for Text Detection”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, pp. 9365–9374.
4. Mingxing Tan, Ruoming Pang, and Quoc V Le. “Efficientdet: Scalable and efficient object detection”. In: arXiv preprint arXiv: 1911.09070 (2019).
5. Piotr Bojanowski et al. “Enriching Word Vectors with Subword Information”. In: Transactions of the Association for Computational Linguistics5 (2017), pp. 135–146.issn: 2307-387X
6. Jiatao Gu et al. “Incorporating Copying Mechanism in Sequence-to-Sequence Learning”. In: CoRRabs/1603.06393 (2016). arXiv: 1603.06393. URL: http://arxiv.org/abs/1603.06393.
7. Steven J. Rennie et al. “Self-critical Sequence Training for Image Captioning”. In: CoRRabs/1612.00563 (2016). arXiv: 1612.00563. URL: http://arxiv.org/abs/1612.00563.
8. Tianyi Zhang et al. “BERTScore: Evaluating Text Generation with BERT”. In: CoRRabs/1904.09675 (2019). arXiv: 1904.09675. URL: http://arxiv.org/abs/1904.09675.

| Copy mechanism included | Transformer Architectures | Test-dev CIDEr Score |
|---|---|---|
| ✗ | *Same* | 76.6 |
| ✗ | *Varying* | 78.0 |
| ✓ | *Same* | 75.1 |
| ✓ | *Varying* | 75.6 |

**Total Ensemble** (90 models, with post-processing) **80.38**

**Open Source**



**Ensemble**

**Pre/Post-process**

Icon credits: Open source: https://icons8.com/icon/65836/open-source Org Chart: https://webstockreview.net/explore/organization-clipart-org-chart/ Paintbrush: https://www.clipartkey.com/view/bwhhbo_transparent-paintbrush-clipart-black-and-white-paint-brush/

# Open Source



## Ensemble



## Pre/Post-process

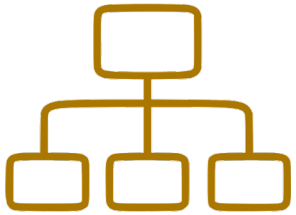
*Leveraging existing implementation for OCR and Object Detection modules made incorporation into pipeline more seamless*


Jeonghun Baek et al. “What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis”. In: International Conference on Computer Vision (ICCV). to appear. 2019. published.

Youngmin Baek et al. “Character Region Awareness for Text Detection”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, pp. 9365–9374.

Mingxing Tan, Ruoming Pang, and Quoc V Le. “Efficientdet: Scalable and efficient object detection”. In: arXiv preprint arXiv: 1911.09070 (2019).

Open Source



**Ensemble**



Pre/Post-process

*Ensembling varying Transformer architectures and combining copy / non-copy mechanism injected caption diversity*

Icon credits: Open source: https://icons8.com/icon/65836/open-source Org Chart: https://webstockreview.net/explore/organization-clipart-org-chart/ Paintbrush: https://www.clipartkey.com/view/bwhhbo_transparent-paintbrush-clipart-black-and-white-paint-brush/

Open Source



Ensemble

**Pre/Post-process**

*Combining de-tokenization, Self Bert scoring, and OCR-maximization provided additional gains*

Icon credits: Open source: https://icons8.com/icon/65836/open-source Org Chart: https://webstockreview.net/explore/organization-clipart-org-chart/ Paintbrush: https://www.clipartkey.com/view/bwhhbo_transparent-paintbrush-clipart-black-and-white-paint-brush/

# Live image captioning with text-to-speech

**Richard Young**

Icon credit: Play button: https://freepngimg.com/png/25623-play-button-transparent-background/icon

# Thank you!