

Third Workshop on Multimodal Artificial Intelligence (MAI 2021)

Online
6 June 2021

ISBN: 978-1-7138-3023-8

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2021) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Multimodal Weighted Fusion of Transformers for Movie Genre Classification</i> Isaac Rodríguez Bribiesca, Adrián Pastor López Monroy and Manuel Montes-y-Gómez	1
<i>On Randomized Classification Layers and Their Implications in Natural Language Generation</i> Gal-Lev Shalev, Gabi Shalev and Joseph Keshet	6
<i>COIN: Conversational Interactive Networks for Emotion Recognition in Conversation</i> Haidong Zhang and Yekun Chai	12
<i>A First Look: Towards Explainable TextVQA Models via Visual and Textual Explanations</i> Varun Nagaraj Rao, Xingjian Zhen, Karen Hovsepian and Mingwei Shen	19
<i>Multi Task Learning based Framework for Multimodal Classification</i> Danting Zeng	30
<i>Validity-Based Sampling and Smoothing Methods for Multiple Reference Image Captioning</i> Shunta Nagasawa, Yotaro Watanabe and Hitoshi Iyatomi	36
<i>Modality-specific Distillation</i> Woojeong Jin, Maziar Sanjabi, Shaoliang Nie, Liang Tan, Xiang Ren and Hamed Firooz	42
<i>Cold Start Problem For Automated Live Video Comments</i> Hao Wu, François Pitie and Gareth Jones	54
<i>¡Qué maravilla! Multimodal Sarcasm Detection in Spanish: a Dataset and a Baseline</i> Khalid Alnajjar and Mika Hämmäläinen	63
<i>A Package for Learning on Tabular and Text Data with Transformers</i> Ken Gu and Akshay Budhkar	69
<i>Semantic Aligned Multi-modal Transformer for Vision-Language Understanding: A Preliminary Study on Visual QA</i> Han Ding, Li Erran Li, Zhiting Hu, Yi Xu, Dilek Hakkani-Tur, Zheng Du and Belinda Zeng	74
<i>GraphVQA: Language-Guided Graph Neural Networks for Graph-based Visual Question Answering</i> Weixin Liang, Yanhao Jiang and Zixuan Liu	79
<i>Learning to Select Question-Relevant Relations for Visual Question Answering</i> Jaewoong Lee, Heejoon Lee, Hwanhee Lee and Kyomin Jung	87