MEG: Multi-Evidence GNN for Multimodal Semantic Forensics

Ekraam Sabir, Ayush Jaiswal, Wael AbdAlmageed, Prem Natarajan
USC Information Sciences Institute, Marina Del Rey, CA, USA
Fake news often involves semantic manipulation of images

Misrepresentation of images with multimodal metadata

Also known as **Image Repurposing** in literature

Image Repurposing involves:

- untampered images
- questionable metadata

Caption: New York is amazing! At Liberty Island.
GPS: 40.6892° N, 74.0446° W
Multimedia Package

• Comprises:
  – Image
  – Related metadata
    • Text
    • Location information

<table>
<thead>
<tr>
<th>Ex Convento at Dorena Bridge completed in 1555</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States, Lane, Oregon, Row River. GPS: 43.737553, -122.883646</td>
</tr>
</tbody>
</table>
Image Repurposing Detection

Reference Dataset

Image + Metadata

Query Image

Query Metadata

Repurposing Detection System

Integrity Score
A novel graph neural network (GNN) based approach to image repurposing detection, that leverages multiple evidences.

Our contributions:

• State-of-the-art performance on three standard benchmarks
• Order-invariant model with respect to the retrieved packages
• Flexible model that can handle arbitrary number of retrievals
The model has a branch for each modality in data and two major modules for each branch:

- Attention Feature Matching module and
- Modality Summary module.

Attention feature matching module compares query and retrieved features.

Modality summary module is a graph neural networks (GNNs). Feature matching embeddings populate the GNN.

The GNN updates all nodes from their adjacent nodes.

A feed-forward network assimilates all modalities for manipulation detection.
Model Architecture
Datasets

• Multimodal Entity Image Repurposing (MEIR) Dataset
  – Packages with Image, text and location modalities.
  – Person, Location and Organization entities manipulated.

• Painters by Numbers
  – Dataset introduced as part of a Kaggle competition.
  – Manipulate Ids of painters.

• Google Landmarks
  – Dataset introduced as part of a Kaggle competition.
  – Manipulate Ids of Locations.

• Metric: AUC
## Results - MEIR

<table>
<thead>
<tr>
<th>Package</th>
<th>SRS</th>
<th>DMT + voting</th>
<th>Ours w/o Cross-modal</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63</td>
<td>0.88</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>0.65</td>
<td>0.87</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>0.66</td>
<td>0.86</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>4</td>
<td>0.67</td>
<td>0.86</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>0.67</td>
<td>0.85</td>
<td>0.90</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Examples

Query

United Kingdom Isle of Wight England Atherfield
GPS: 50.611785, -1.363763

Draft Horse Town 2014

Canada Alberta Alberta Calgary
GPS: 51.036077, -114.054468

Draft Horse Town 2014

Retrieved Packages

Hockwold Country Fair
20132nd June ... Samba Dancers Morris Dancers

United Kingdom Isle of Wight England Atherfield
GPS: 52.463838, -0.558081

Canada Alberta Alberta Calgary
GPS: 51.036077, -114.054468

United Kingdom Isle of Wight England Atherfield

The Rotunda of Galerius is 125m northeast of the Arch of Galerius It is now the GM, ... dedicated UNESCO World Heritage Site ...

Greece Central Macedonia Macedonia and Thrace Thessaloniki
GPS: 40.633276, 22.952864

The Rotunda of Galerius is 125m ... It is now the Greek Orthodox Church of Agios ... dedicated UNESCO World Heritage Site ...

Greece Central Macedonia Macedonia and Thrace Thessaloniki
GPS: 40.633276, 22.952864

The Church of St Adalbert or Church of St Wojciech ... and gave his sermons there Wikipedia.org]

Poland Krakow Lesser Poland Krakow
GPS: 50.060927, 19.937650

The German Fountain is a gazebo styled fountain in ... and dome's interior is covered with golden mosaics Wikipedia.org

Turkey Kartal Istanbul Istanbul
GPS: 41.006944, 28.976667
Results – Painters by Numbers

<table>
<thead>
<tr>
<th>Package</th>
<th>SRS</th>
<th>DMT + voting</th>
<th>Ours w/o Cross-modal</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>0.75</td>
<td>0.69</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.80</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td>0.74</td>
<td>0.82</td>
<td>0.79</td>
<td>0.83</td>
</tr>
<tr>
<td>4</td>
<td>0.76</td>
<td>0.83</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>5</td>
<td>0.77</td>
<td>0.83</td>
<td>0.81</td>
<td>0.86</td>
</tr>
</tbody>
</table>
## Results – Google Landmarks

<table>
<thead>
<tr>
<th>Package</th>
<th>SRS</th>
<th>DMT + voting</th>
<th>Ours w/o Cross-modal</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.89</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>3</td>
<td>0.92</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>
## Results – Order Invariance

<table>
<thead>
<tr>
<th>Packages</th>
<th>GRU Before</th>
<th>GRU After</th>
<th>LSTM Before</th>
<th>LSTM After</th>
<th>Ours w/o cross-modal Before</th>
<th>Ours w/o cross-modal After</th>
<th>Ours Before</th>
<th>Ours After</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>0.88</td>
<td>0.90</td>
<td>0.89</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>0.91</td>
<td>0.85</td>
<td>0.91</td>
<td>0.88</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>4</td>
<td>0.91</td>
<td>0.84</td>
<td>0.91</td>
<td>0.88</td>
<td>0.90</td>
<td>0.90</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>0.91</td>
<td>0.83</td>
<td>0.91</td>
<td>0.87</td>
<td>0.90</td>
<td>0.90</td>
<td>0.92</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Thank You!