Unsupervised Alignment Network
Hala Lamdouar, WeiDi Xie, Andrew Zisserman.

Visual Geometry Group VGG
Department Of Engineering Science
University of Oxford

Motivation:

Image alignment is a classical task that involves finding correspondeces and inferring geometric transformation that maps a given source image to a target image.

- Current methods lack robustness and fail to align the complex and ambiguous cases.

Contributions:

- Unsupervised robust alignment network that emphasises relevant features and handles outliers.
- Burstiness module that penalises the repetitive regions or the textures within the image.

Self-Burstiness:

Consider an image \( I \in \mathbb{R}^{M \times N} \) and its feature representation \( F_I \in \mathbb{R}^{M \times N \times D} \).

For \((i,j) \in I\) \[ U_{i,j}^l = \text{softmax}(B_r(f_{i,j})) \]

- Burstiness-based matching score for each \((i,j) \in I_{src}, (k,l) \in I_{tgt}\) \[ s_{i,j,k,l} = u_{i,j} \cdot u_{k,l} \]  

RANSAC-like inlier mask:

For each \((i,j) \in I_{src}, (k,l) \in I_{tgt}\)
\[ m_{i,j,k,l} = \begin{cases} 1 & \text{if } \| (i,j) - T^{-1}(k,l) \|_2 < \ell \\ 0 & \text{otherwise} \end{cases} \]

Model:

End-to-end alignment network including:
- A pre-trained feature extraction convolutional neural network (VGG-16, ResNet-101)
- A burstiness module (downweights irrelevant features)
- 2D geometric transform generation (convolutional regression network)
- A differentiable RANSAC-like inlier count (inlier mask)

Loss function:

\[ L_{\text{inliers}} = - \sum_{(i,j) \in I_{src}} \sum_{(k,l) \in I_{tgt}} s_{i,j,k,l} \cdot \ell_{src,tgt}(i,j) \cdot m_{i,j,k,l} \]

Challenging dataset:

- 2 different transformations for the background (PASCAL VOC2011) and the foreground (3D rendered chairs)
- Cropping with respect to the maximal inscribed axis-aligned rectangle

Results:

We use the PCK measure (percentage of correct points) i.e. the number of points where:
\[ \| T_0(i,j) - T_{tgt}(i,j) \|_2 < \ell \]

On PF-PASCAL:

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall (%)</th>
<th>AP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG-16</td>
<td>43.65</td>
<td>33.59</td>
</tr>
<tr>
<td>ResNet-101</td>
<td>43.81</td>
<td>33.60</td>
</tr>
<tr>
<td>VGG-16 + RANSAC</td>
<td>48.06</td>
<td>34.75</td>
</tr>
<tr>
<td>ResNet-101 + RANSAC</td>
<td>48.20</td>
<td>34.80</td>
</tr>
</tbody>
</table>

References: