**Abstract:** We tackle the problem of visual search under resource constraints. Existing systems use the same embedding model to compute embeddings for the query and gallery images necessitating a hard accuracy-efficiency tradeoff. We mitigate the tradeoff by proposing a heterogeneous visual search (HVS) system leading to 80 fold and 23 fold cost reduction for challenging retrieval problems on fashion (DeepFashion2) and face (IJBB-C) images. This is achieved with marginal 0.3% and 1.6% drop in accuracy. Key to developing an HVS system is to ensure representational compatibility between the query and gallery embedding models.

**Compatibility Criterion**

Given models $(\phi_Q, \phi_G)$

Such that $\text{Size}(\phi_Q) \ll \text{Size}(\phi_G)$

$M(\phi_Q, \phi_G) < M(\phi_Q, \phi_G) < M(\phi_Q, \phi_G)$

$M$ is a metric such as Top-k.

**Compatibility through weights**

**Compatibility through architecture (CMP-NAS)**

**Key Results**

<table>
<thead>
<tr>
<th>Model</th>
<th>Homogeneous Acc.</th>
<th>Heterogeneous Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>65.2%</td>
<td>86.7%</td>
</tr>
<tr>
<td>Baseline</td>
<td>56.0%</td>
<td>77.1%</td>
</tr>
</tbody>
</table>

**Fig. 1** A homogeneous visual search system with a shared embedding model.

**Fig. 2** A heterogeneous visual search system with decoupled embedding models.

**Fig. 3** Accuracy (top) and efficiency (bottom) of a heterogeneous visual search system.

**Fig. 4** Weight level compatibility.

1. Sample architecture $a = A/(X)[2]$.
3. Minimize reward $a = \arg\max_a \mathbb{E}[P(a)]$

**Fig. 5** Architecture level compatibility.

**Fig. 6** Face Retrieval on IJB-C.

**Fig. 7** Fashion Retrieval on DeepFashion-2.

**Fig. 8** Comparison with traditional NAS on the Faces (left) and Fashion (right) datasets.

---
