Self-Supervised Contrastive Learning: SimCLR
What is Contrastive Learning?

The quest for a representation $f : X \rightarrow Z$ where:

- Similar data (positive pairs) $(x, y) \sim p^+$ are close
- Dissimilar data (negative pairs) $(x, y) \sim p^-$ are distant

Loss archetype (‘siamese loss’):

$$\mathbb{E}_{p^+} [d(f(x), f(y))] - \mathbb{E}_{p^-} [d(f(x), f(y))]$$
Two Flavours of Contrastive Learning

**Supervised**: positive pairs have **same class**. Dates way back.

**Un/Self -Supervised**: positive pairs are related by **data augmentation**.

Hadsell, Chopra, *LeCun*,
Comparisons with other un/self-supervised representation learning frameworks, ex. Autoencoders.
Details on SimCLR: Data Augmentation

(a) Original  (b) Crop and resize  (c) Crop, resize (and flip)  (d) Color distort. (drop)  (e) Color distort. (jitter)

(f) Rotate \{90^\circ, 180^\circ, 270^\circ\}  (g) Cutout  (h) Gaussian noise  (i) Gaussian blur  (j) Sobel filtering

...And their compositions!
Details on SimCLR: Loss

Softmax’d cosine distance. Everything in the **big batches** is dissimilar. For a similar pair within a batch, the loss is:

\[- \log \frac{e^{\cos(z_i, z_j) / \tau}}{\sum_{k \neq i} e^{\cos(z_i, z_k) / \tau}}\]
The representation one should use is actual an intermediate one. In other words, a **projection head** $g$ is attached on top of the model.
Discussion Points

Why composing transformations?
Why attaching a (complex) projection head?
Weak Points and How to Solve Them

- Memory cost of batch gradients.
  
  A solution: MoCO. He et al., *Momentum Contrast for Unsupervised Visual Representation Learning*
  
  It is incorporated in SimCLRv2: Chen et al., *Big Self-Supervised Models are Strong Semi-Supervised Learners*

- Too much invariance, even with projection head.
  
  A solution: LooC. Xiao et al., *What Should not be Contrastive in Contrastive Learning*
Results: Comparisons

Representations are evaluated by training a linear classifier on top.
Results: Transfer Learning

Models are pretrained on ImageNet.

<table>
<thead>
<tr>
<th></th>
<th>Food</th>
<th>CIFAR10</th>
<th>CIFAR100</th>
<th>Birds</th>
<th>SUN397</th>
<th>Cars</th>
<th>Aircraft</th>
<th>VOC2007</th>
<th>DTD</th>
<th>Pets</th>
<th>Caltech-101</th>
<th>Flowers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linear evaluation:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SimCLR (ours)</td>
<td>76.9</td>
<td>95.3</td>
<td>80.2</td>
<td>48.4</td>
<td>65.9</td>
<td>60.0</td>
<td>61.2</td>
<td>84.2</td>
<td>78.9</td>
<td>89.2</td>
<td>93.9</td>
<td>95.0</td>
</tr>
<tr>
<td>Supervised</td>
<td>75.2</td>
<td>95.7</td>
<td>81.2</td>
<td>56.4</td>
<td>64.9</td>
<td>68.8</td>
<td>63.8</td>
<td>83.8</td>
<td>78.7</td>
<td>92.3</td>
<td>94.1</td>
<td>94.2</td>
</tr>
<tr>
<td><strong>Fine-tuned:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SimCLR (ours)</td>
<td>89.4</td>
<td>98.6</td>
<td>89.0</td>
<td>78.2</td>
<td>68.1</td>
<td>92.1</td>
<td>87.0</td>
<td>86.6</td>
<td>77.8</td>
<td>92.1</td>
<td>94.1</td>
<td>97.6</td>
</tr>
<tr>
<td>Supervised</td>
<td>88.7</td>
<td>98.3</td>
<td>88.7</td>
<td>77.8</td>
<td>67.0</td>
<td>91.4</td>
<td>88.0</td>
<td>86.5</td>
<td>78.8</td>
<td>93.2</td>
<td>94.2</td>
<td>98.0</td>
</tr>
<tr>
<td>Random init</td>
<td>88.3</td>
<td>96.0</td>
<td>81.9</td>
<td>77.0</td>
<td>53.7</td>
<td>91.3</td>
<td>84.8</td>
<td>69.4</td>
<td>64.1</td>
<td>82.7</td>
<td>72.5</td>
<td>92.5</td>
</tr>
</tbody>
</table>
Free Discussion
Opinions About the Paper

- Making contrastive learning self-supervised: elegantly simple and useful idea.

- Great results.

- The paper contains a lot of ablations. Maybe too many?
Tack!