Motivation

Aim: universal representation.

Goal: a deep distributed representation that can exploit synergies between tasks, find coherent solutions, and be easily extensible to new tasks.

Contributions:
1. We show that a single shared network performs better than individual ones when trained on multiple tasks, such as image classification and object detection.
2. We introduce a ‘integration space’ which is iteratively updated as new information is discovered by solving individual tasks; this couples the tasks together, in a recurrent manner.
3. We show that the latter idea further improves the results.

OVERVIEW

MULTINET

Task $\alpha$ is specified by the triplet: $T^\alpha = (X^\alpha, \phi^\alpha_{\text{enc}}, \psi^\alpha_{\text{dec}})$

- Encoder function maps the label $x^\alpha$ to a vector $r^\alpha \in \mathcal{R}^\alpha$
  $$r^\alpha = \phi^\alpha_{\text{enc}}(x^\alpha)$$
- Decoder function maps a common representation $h$ to the label $x^\alpha$
  $$x^\alpha = \psi^\alpha_{\text{dec}}(h)$$
- Integrator function maps $r^0, r^1, \ldots, r^K$ to the common subspace $h$
  $$h_{t+1} = \Gamma(h_t, r^0_t, r^1_t, \ldots, r^K_t)$$

$t = 0$, ordinary multi-task prediction [1]

- Shared representation $h$ is set to $h_0 = \phi^\alpha_{\text{enc}}(x^0) = \Gamma(x^0, * , * , * , * )$
- Prediction for task $\alpha$ is $x^\alpha_t = \psi^\alpha_{\text{dec}}(h_t) = (\psi^\alpha_{\text{dec}} \circ \phi^\alpha_{\text{enc}})(x^\alpha_t)$

$t > 0$, iterative updates

- Shared representation $h_{t+1} = \Gamma(h_t, r^0_t, r^1_t, \ldots, r^K_t)$ using $r^\alpha_t = \phi^\alpha_{\text{enc}}(x^\alpha_t)$
- Labels are predicted again using $x^\alpha_{t+1} = \psi^\alpha_{\text{dec}}(h_{t+1})$

RESULTS

Tasks/Datasets
- PASCAL VOC 2007 (image classification, object detection)
- PASCAL VOC 2010 (image classification, object, part detection)

Network
- Object and part detection with Fast-RCNN [2]

Encoder functions
- Classification $\forall u, v, c : \ r^{cls}_{uv} = x^{cls}_u$
- Detection $\forall u, v, c : \ r^{det}_{uv} = \max \{ r^{det}_{uv} \in \mathcal{R}^{det} \}_{(u,v) \in b_m} \cup \{0\}$

Update functions
- $h_t = \Gamma(h_{t-1}, r^{img}_{t}, r^{cls}_{t}, r^{det}_{t}, r^{part}_{t}) = \text{stack}(r^{img}_{t}, r^{cls}_{t}, r^{det}_{t}, r^{part}_{t})$
- $h_t = \Gamma(h_{t-1}, r^{img}_{t}, r^{cls}_{t}, r^{det}_{t}, r^{part}_{t}) = \text{ReLU}(A \text{stack}(h_{t-1}, r^{img}_{t}, r^{cls}_{t}, r^{det}_{t}, r^{part}_{t}))$

This work was supported by ERC Starting Grant Integrated and Detailed Image Understanding (EP/L024683/1)