2. End-to-End Multi-Task Learning with Attention

- MTAN = a single shared network (task-shared features)+ a soft-attention module (task-specific features)

- Training multiple tasks' two key challenges

  - Network Architecture (how to share):

    - generalisable representation (to avoid over-fitting)

    - features tailored to each task (to avoid under-fitting)

    - $\Rightarrow$ enables both task-shared and task-specific features to be learned automatically, and consequently

  - Loss Function (how to balance task):

    - A multi-task loss function, which weights the relative contributions of each task, should enable learning of all tasks with equal importance.

    - $\Rightarrow$ learns an inherent robustness to the choice of loss weighting scheme.

- each soft attention mask automatically determines the importance of the shared features for the respective task.

- can be built on any feed-forward neural network

  - SegNet (an encoder-decoder network) w/ CityScapes dataset, NYUv2

  - Wide Residual Network w/ Visual Decathlon Challenge

  - Results

    - outperforms several baselines, competitive with SOTA, more parameter efficient

    - Dynamic Weight Average: adapts the task weighting over time by considering the rate of change of the loss for each task.