

LEGEND

Solid arrow:
Forward dataflow



Dashed arrow:
Fusing



Black symbols:
Forward

$$N = \mathbf{H} \cdot \mathbf{W}$$

Red symbols:
Backward

$$\partial \mathbf{W} = \mathbf{H}^T \cdot \partial N$$

Green circle:
Embedding



Violet circle:
Parameter



Grey circle:
Virtual tensor



**Abbreviation
for gradient:**

$$\partial W := \partial \mathcal{L} / \partial W$$

**Partial sum
of gradient:**

$$\partial H = \sum_i (\partial H)_i$$