The figure illustrates the overall network architecture. The top panel shows the main pipeline: input $\mathbf{\Phi}^T\mathbf{y}$ of size $1 \times H \times W$ passing through a ResBlock producing $\mathbf{x}_K$ of size $3 \times H \times W$ and $\mathbf{z}_K$ of size $(C-3) \times H \times W$, then a sequence of DMP steps: $\mathbf{x}_{K-1}, \dots, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_1$ and $\mathbf{z}_{K-1}, \dots, \mathbf{z}_t, \mathbf{z}_{t-1}, \dots, \mathbf{z}_1$, producing $\mathbf{x}_0$ of size $3 \times H \times W$ and $\mathbf{z}_0$ of size $(C-3) \times H \times W$, and finally a ResBlock producing $\hat{\mathbf{x}}$ of size $1 \times H \times W$.

The middle panel shows the DMP step. Inputs $\mathbf{x}_t$ and $\mathbf{z}_t$. The Gradient Descent block (3 to 3) produces $\mathbf{s}_t$, which is channel-wise concatenated with $\mathbf{z}_t$ to form $\{\mathbf{s}_t, \mathbf{z}_t\}$. This passes through a ResBlock (C to C) producing $\{\mathbf{r}_t, \mathbf{z}_{t-1}\}$, which is channel-wise split into $\mathbf{r}_t$ and $\mathbf{z}_{t-1}$. Then $p_\theta(\mathbf{x}_{t-1}|\mathbf{r}_t)$ (3 to 3) produces $\mathbf{x}_{t-1}$.

The bottom panel shows building blocks. ResBlock: Conv (C_{in}), Res (32), Res (32), Res (32), Res (32), Conv (32 to C_{out}). Res: Conv (32), ReLU (32), Conv (32), element-wise addition (32). Legend: Ⓢ channel-wise splitting; Ⓒ channel-wise concatenation; ⊕ element-wise addition; Conv: 3×3 convolution with the stride of 1; C_i and C_o represent the number of input and output channels, respectively.