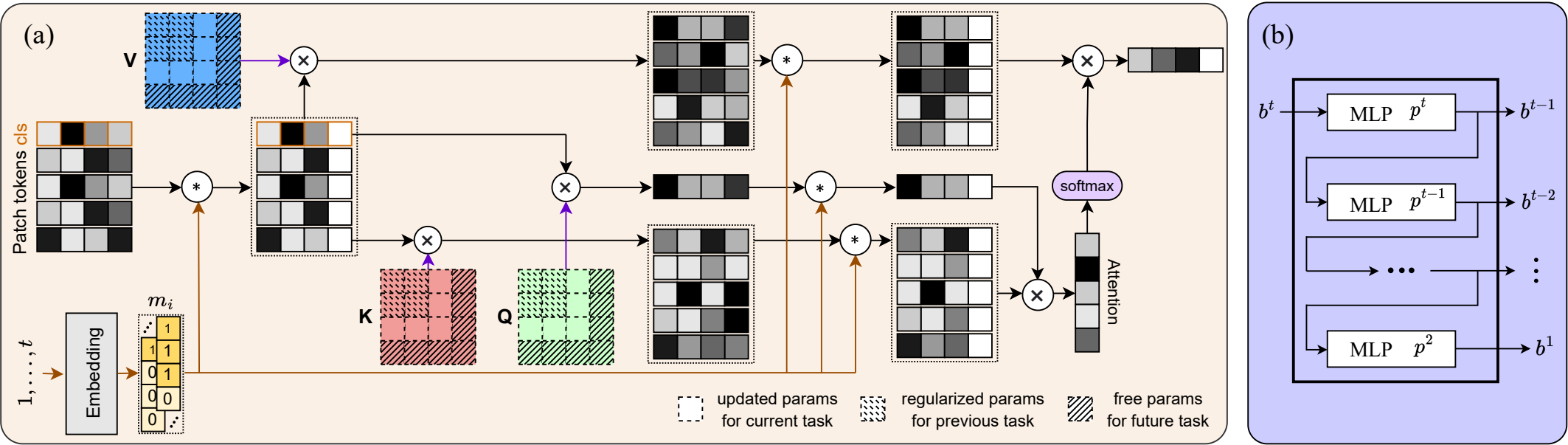
(a)
V
Patch tokens cls
K
Q
m_i
$1, \ldots, t$
Embedding
softmax
Attention
updated params for current task
regularized params for previous task
free params for future task
(b)
b^t
MLP p^t
b^{t-1}
MLP p^{t-1}
b^{t-2}
MLP p^2
b^1