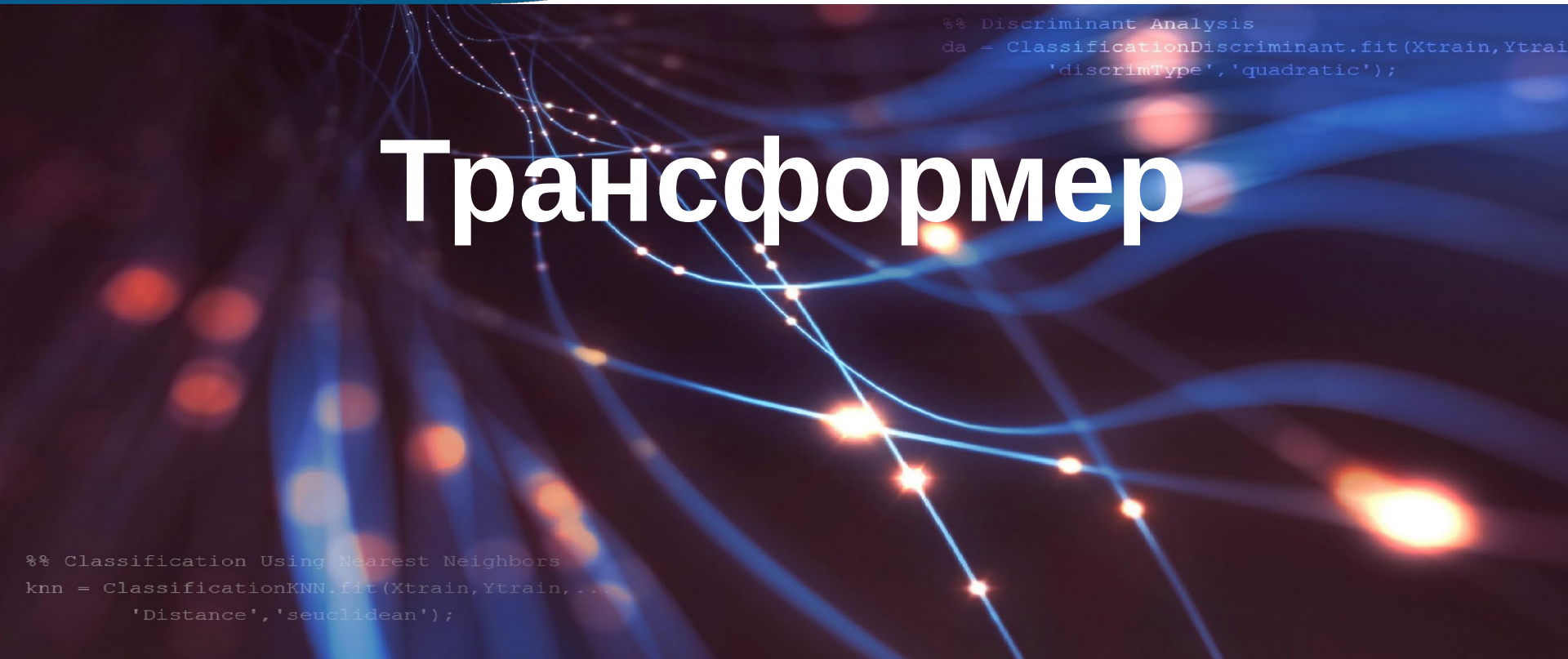
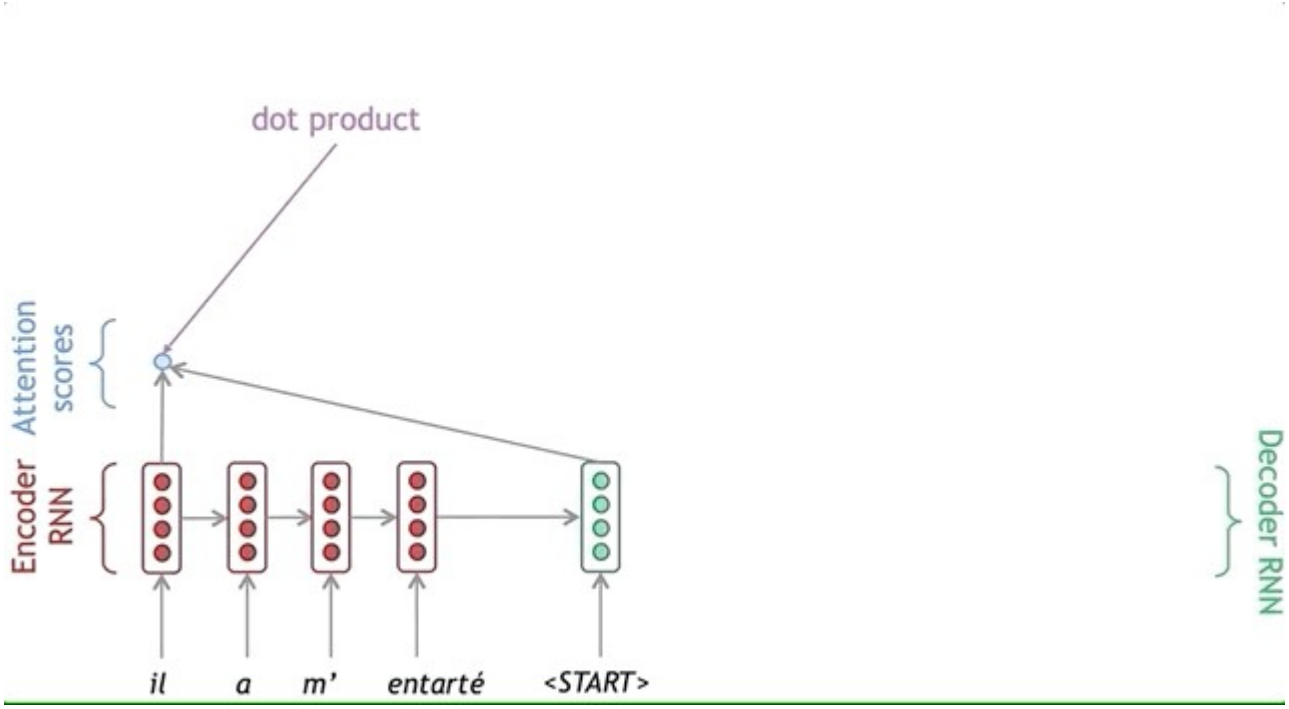
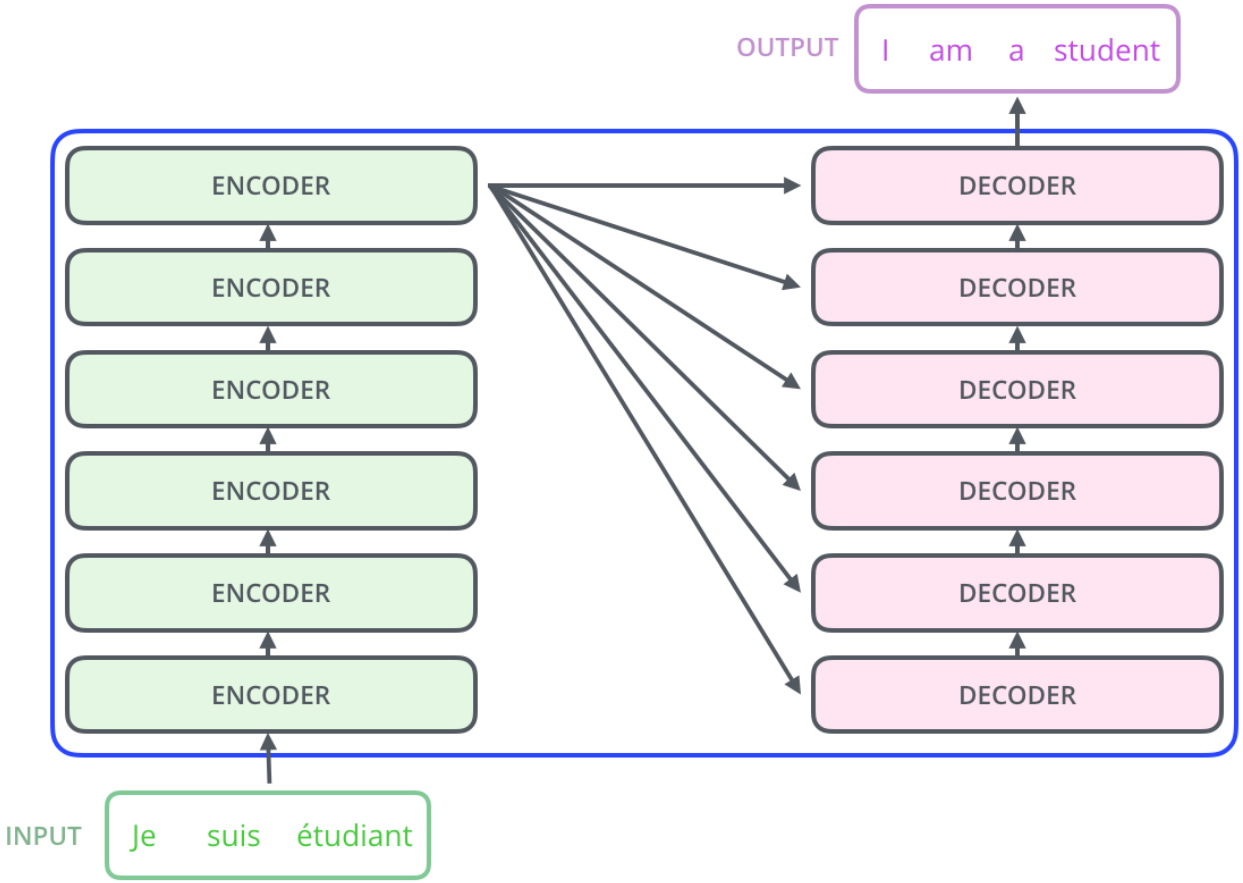
экспонента
Центр Инженерных Технологий
и Моделирования

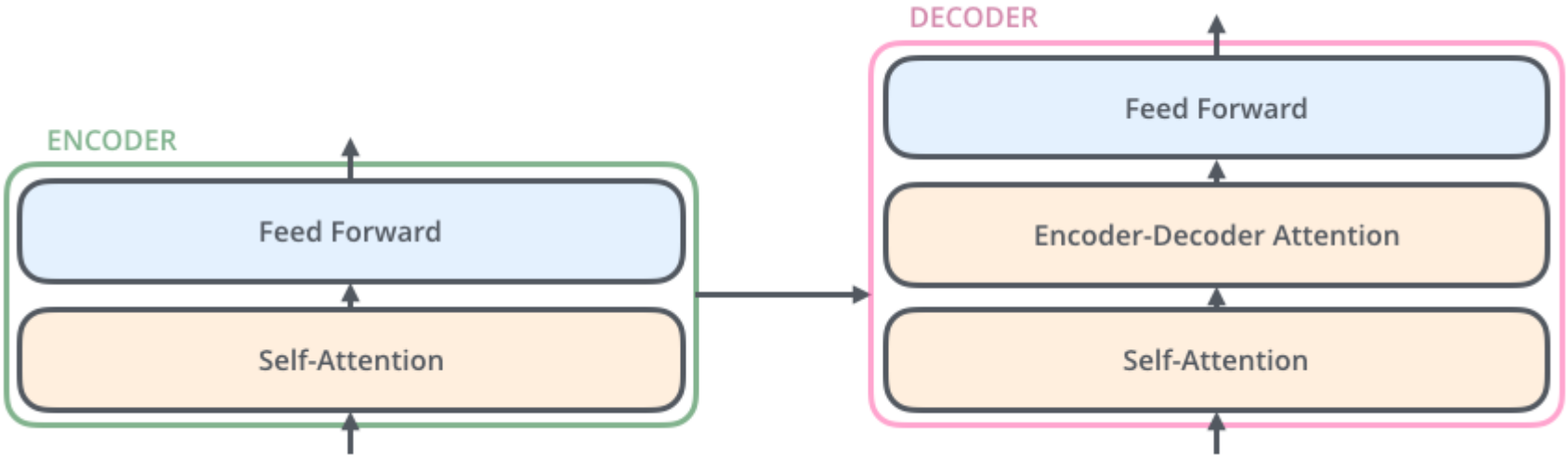
MathWorks®

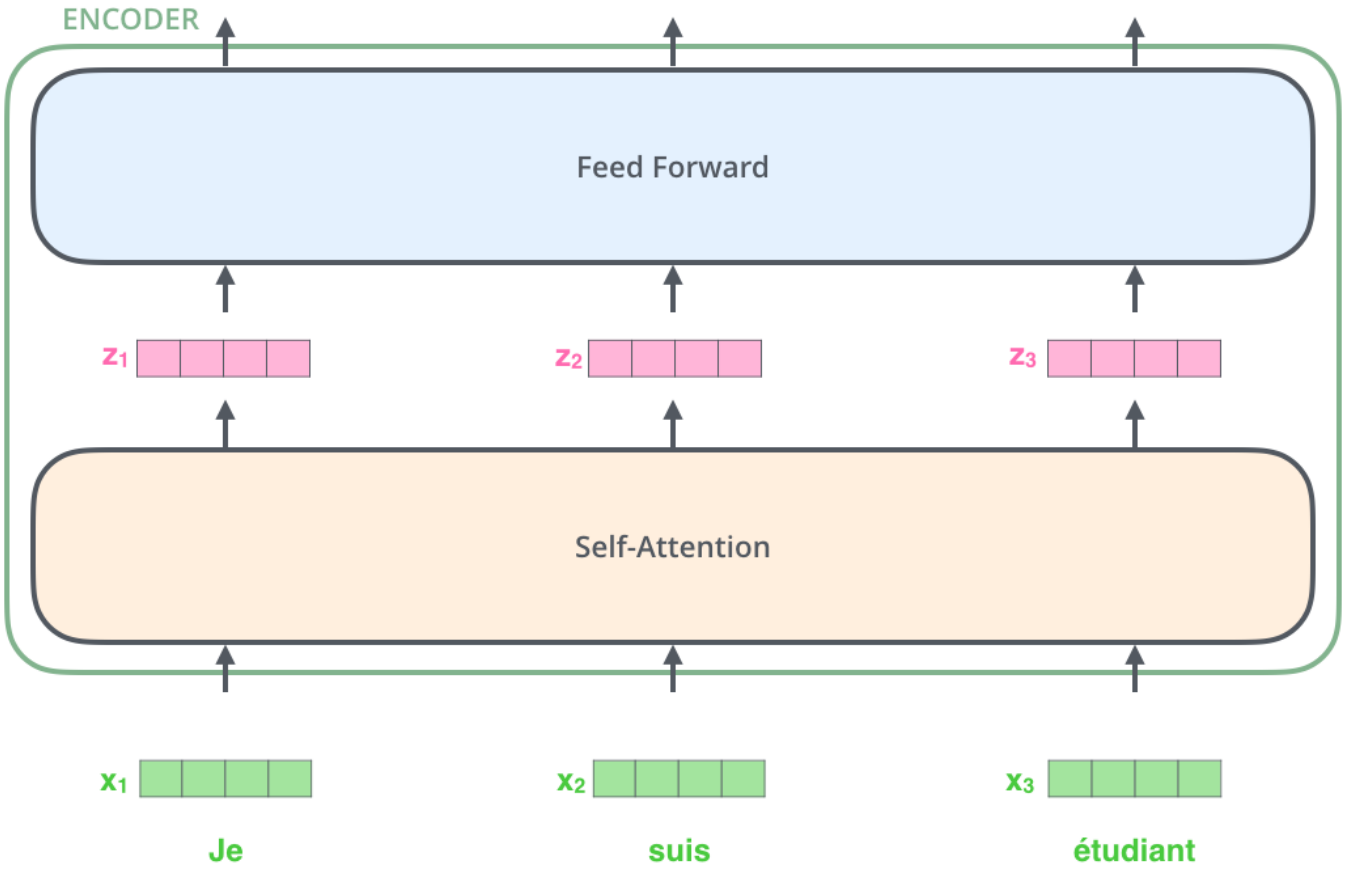


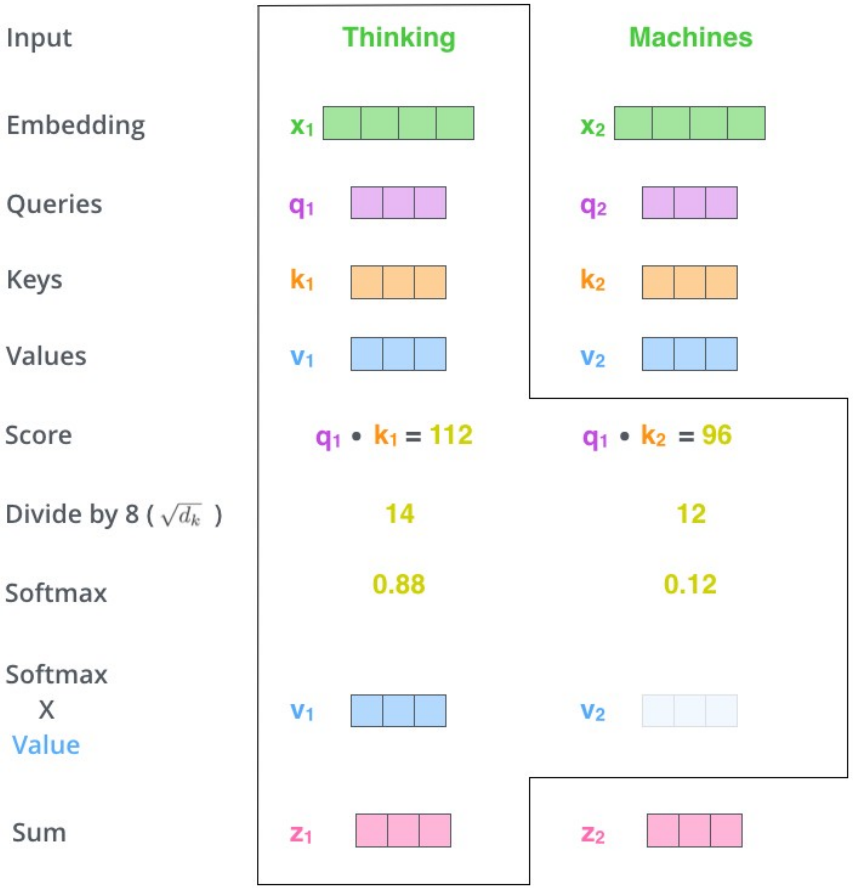
Трансформер

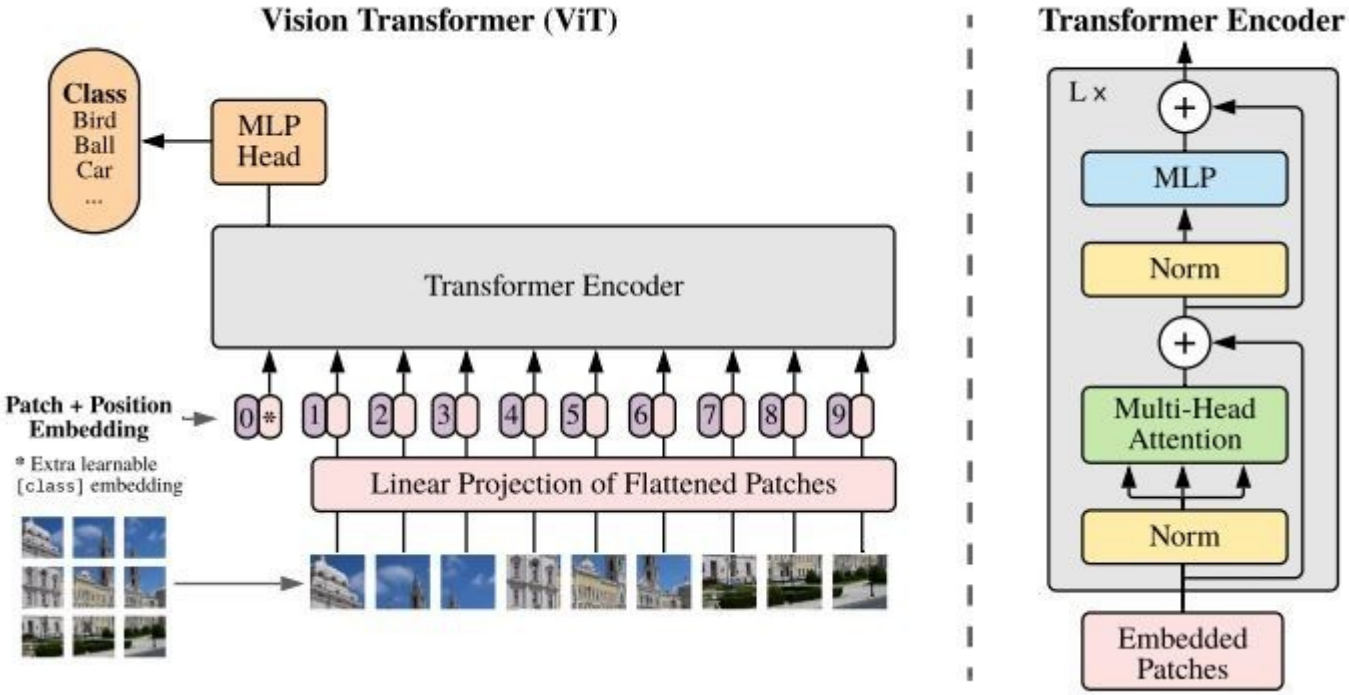

dot product
Attention scores
Encoder RNN
Decoder RNN
il
a
m'
entarté
<START>

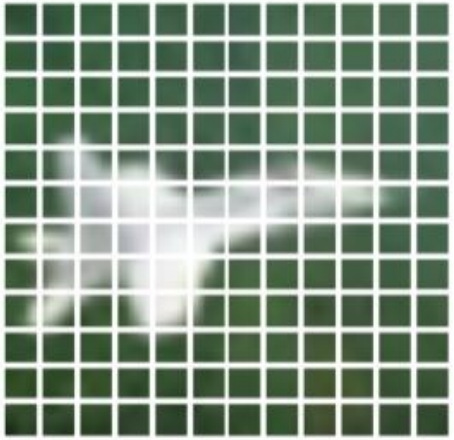

OUTPUT
I am a student
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
DECODER
DECODER
DECODER
DECODER
DECODER
DECODER
INPUT
Je suis étudiant

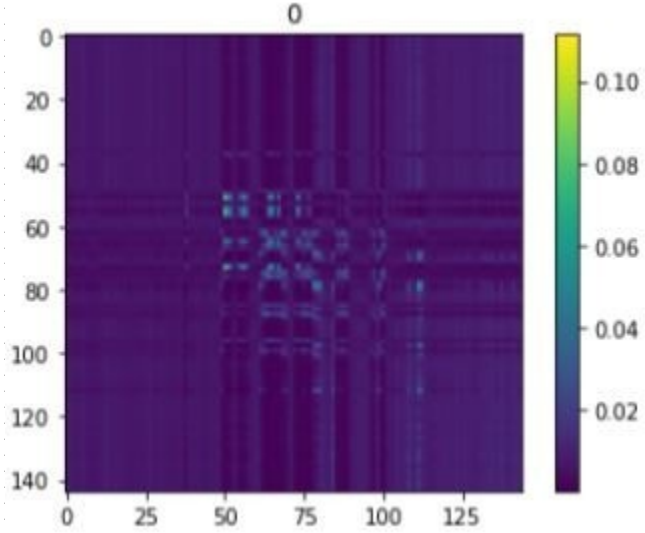

ENCODER
Feed Forward
Self-Attention
DECODER
Feed Forward
Encoder-Decoder Attention
Self-Attention

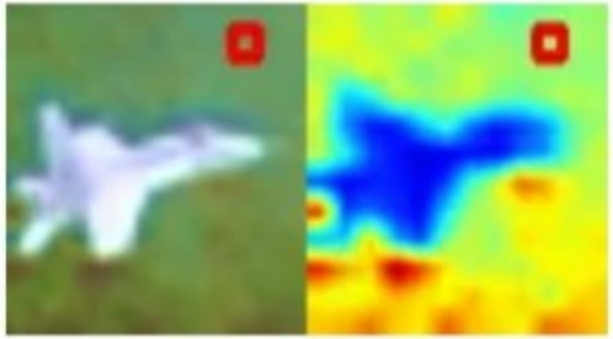

ENCODER
Feed Forward
z1
z2
z3
Self-Attention
x1
x2
x3
Je
suis
étudiant

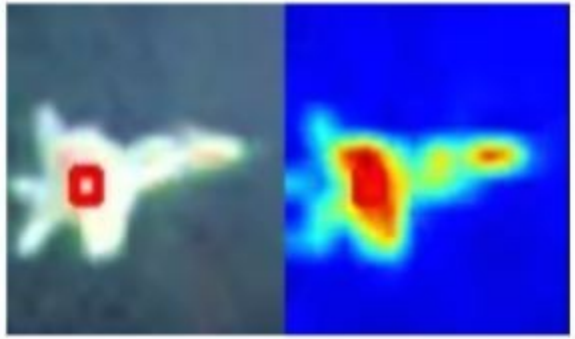

Input
Thinking
Machines
Embedding
x1
x2
Queries
q1
q2
Keys
k1
k2
Values
v1
v2
Score
q1 • k1 = 112
q1 • k2 = 96
Divide by 8 (√dk)
14
12
Softmax
0.88
0.12
Softmax X Value
v1
v2
Sum
z1
z2

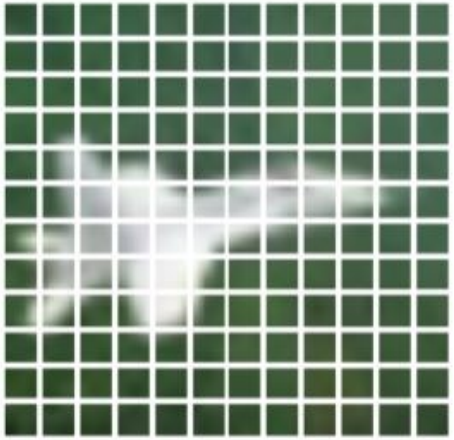

Vision Transformer (ViT)
Class
Bird
Ball
Car
...
MLP Head
Transformer Encoder
Patch + Position Embedding
* Extra learnable [class] embedding
0 * 1 2 3 4 5 6 7 8 9
Linear Projection of Flattened Patches
Transformer Encoder
L x
MLP
Norm
Multi-Head Attention
Norm
Embedded Patches

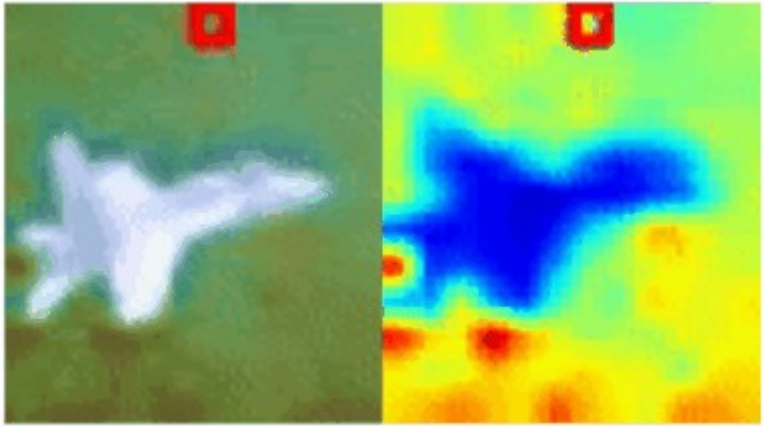



0
0
20
40
60
80
100
120
140
0
25
50
75
100
125
0.10
0.08
0.06
0.04
0.02

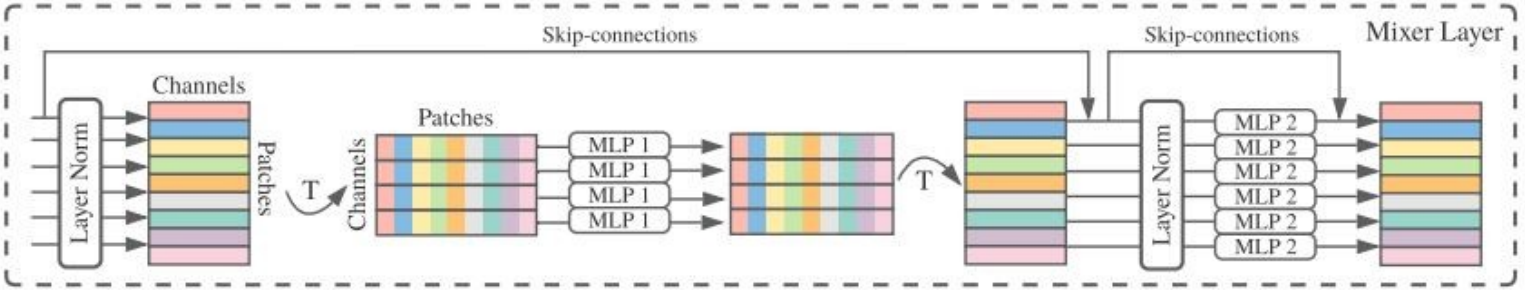


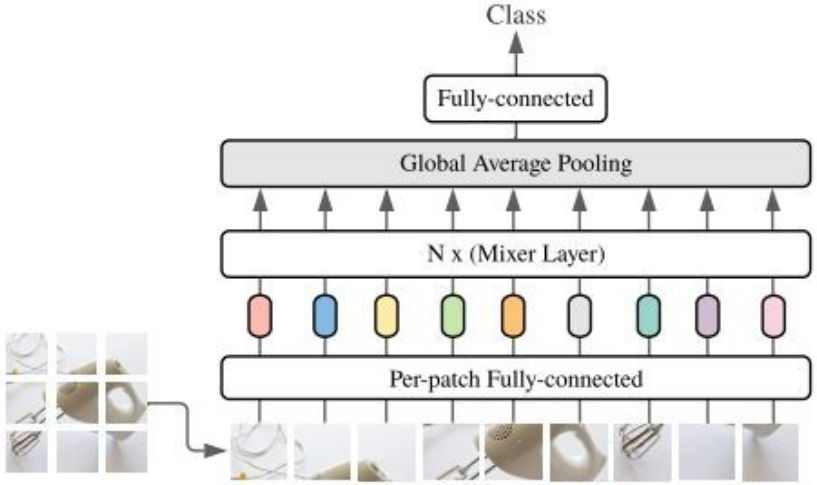


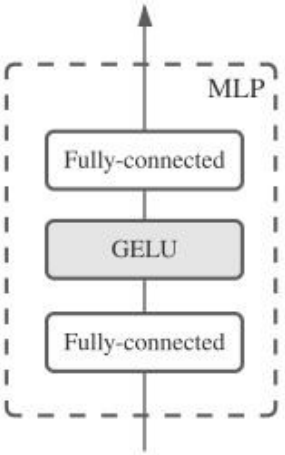




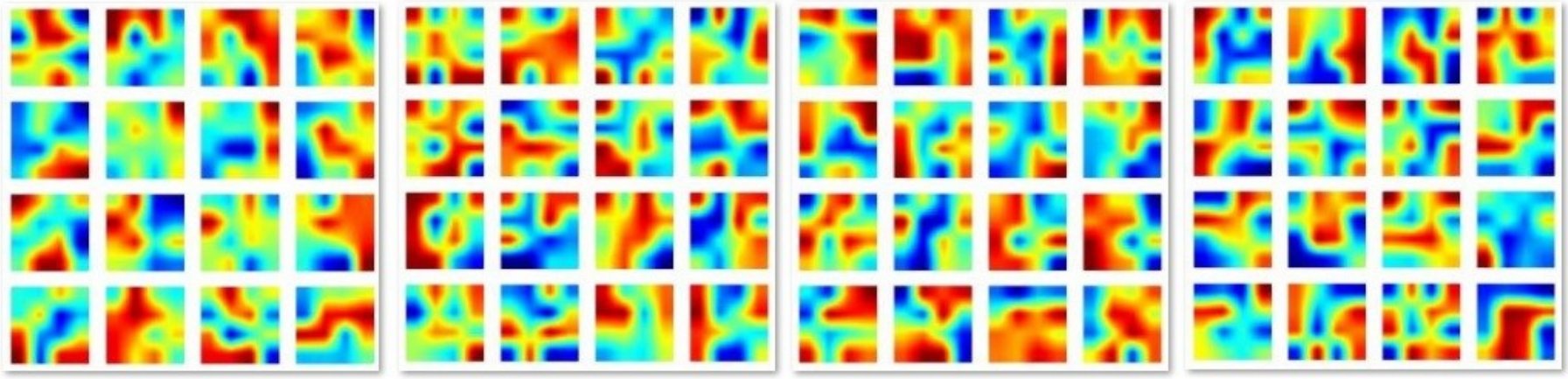
"MLP mixer


Skip-connections
Skip-connections
Mixer Layer
Channels
Layer Norm
Patches
T
Patches
Channels
MLP 1
MLP 1
MLP 1
MLP 1
T
Layer Norm
MLP 2
MLP 2
MLP 2
MLP 2
MLP 2
MLP 2

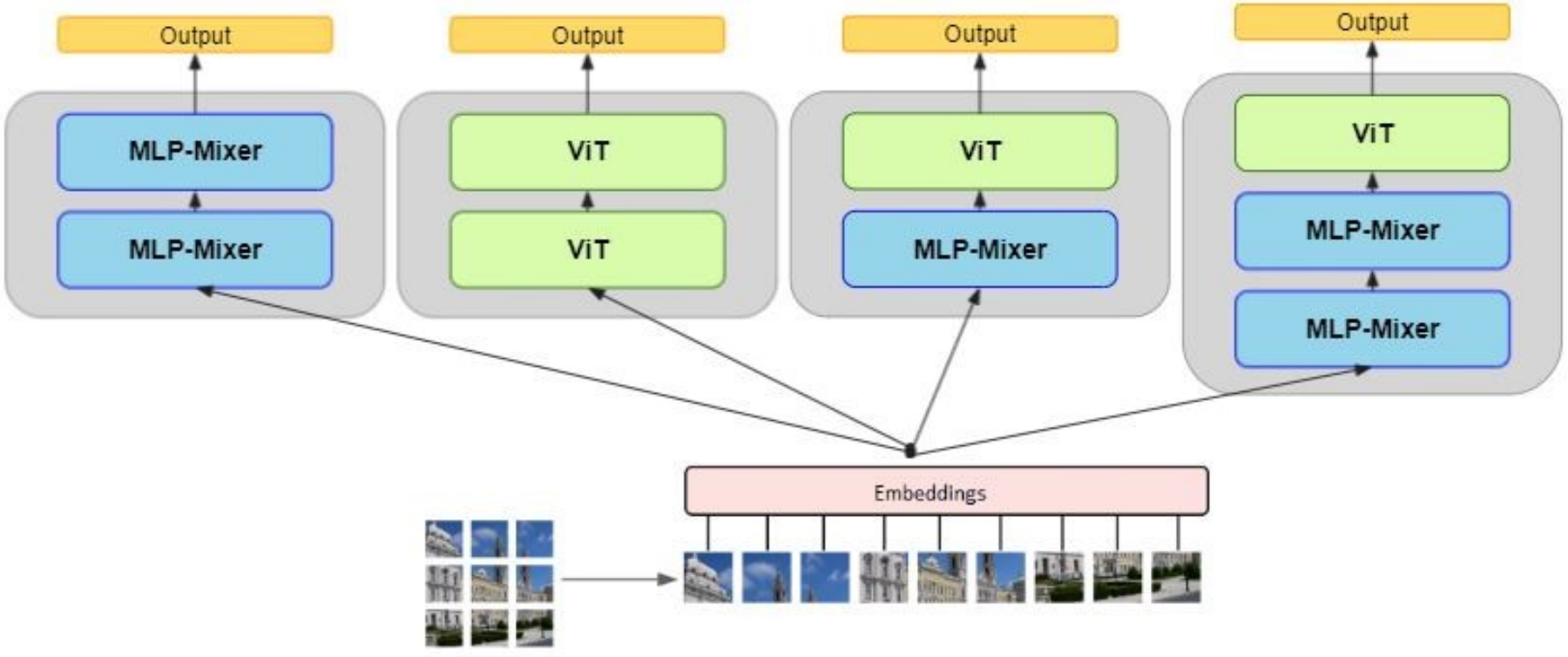

Class
Fully-connected
Global Average Pooling
N x (Mixer Layer)
Per-patch Fully-connected


MLP
Fully-connected
GELU
Fully-connected

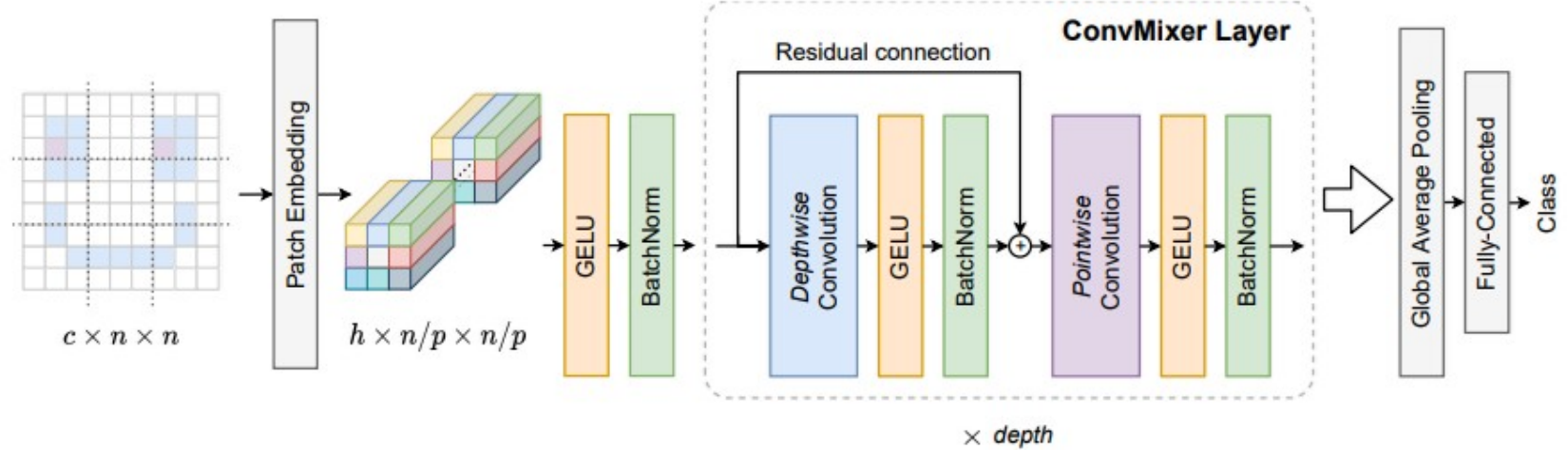
"MLP mixer



Hybrid MLP-Mixer and ViT



CovMixer




```
def ConvMixr(h,d,k,p,n):
 S,C,A=Sequential,Conv2d,lambda x:S(x,GELU(),BatchNorm2d(h))
 R=type('',(S,),{'forward':lambda s,x:s[0](x)+x})
 return S(A(C(3,h,p,p)),*[S(R(A(C(h,h,k,groups=h,padding=k//2))),A(C(h,h,1))) for i
  ↪  in range(d)],AdaptiveAvgPool2d((1,1)),Flatten(),Linear(h,n))
```

Hybrid MLP-Mixer and CovMixer