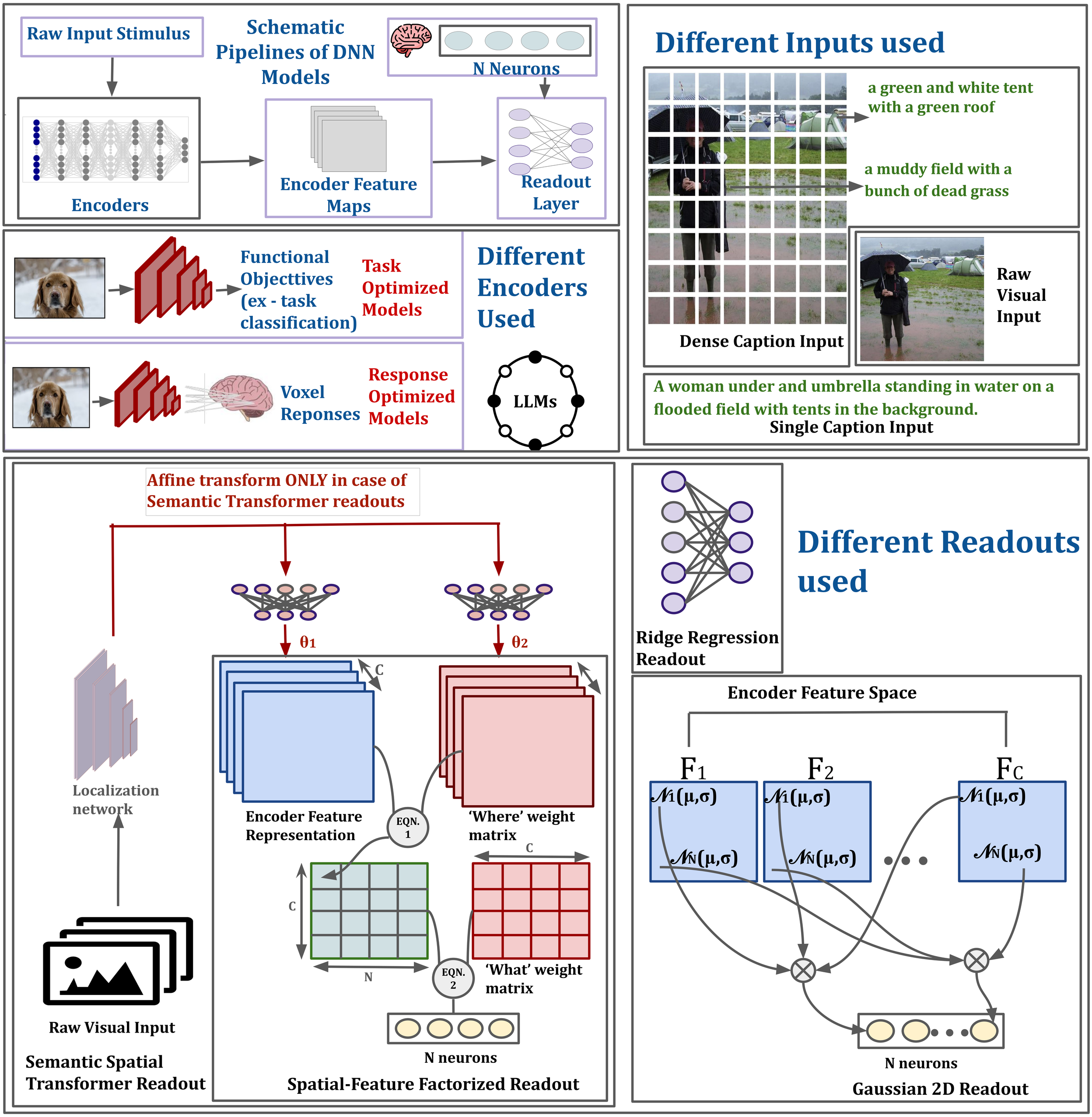
Raw Input Stimulus
Schematic Pipelines of DNN Models
N Neurons
Encoders
Encoder Feature Maps
Readout Layer
Functional Objecttives (ex - task classification)
Task Optimized Models
Voxel Reponses
Response Optimized Models
Different Encoders Used
LLMs
Different Inputs used
a green and white tent with a green roof
a muddy field with a bunch of dead grass
Dense Caption Input
Raw Visual Input
A woman under and umbrella standing in water on a flooded field with tents in the background.
Single Caption Input
Affine transform ONLY in case of Semantic Transformer readouts
θ1
θ2
C
Localization network
Encoder Feature Representation
EQN. 1
'Where' weight matrix
C
C
N
EQN. 2
'What' weight matrix
Raw Visual Input
N neurons
Semantic Spatial Transformer Readout
Spatial-Feature Factorized Readout
Ridge Regression Readout
Different Readouts used
Encoder Feature Space
F1
F2
FC
𝒩1(μ,σ)
𝒩N(μ,σ)
N neurons
Gaussian 2D Readout