The 7th International Conference on Development and Learning and Epigenetic Robotics,
Instituto Superior Tecnico, Lisbon, Portugal, 20th Sep, 2017

Predictive Coding-based Deep Dynamic Neural Network for Visuomotor Learning

Jungsik Hwang[1,2], Jinhyung Kim[1], Ahmadreza Ahmadi[1,2], Minkyu Choi[2] and Jun Tani[2]

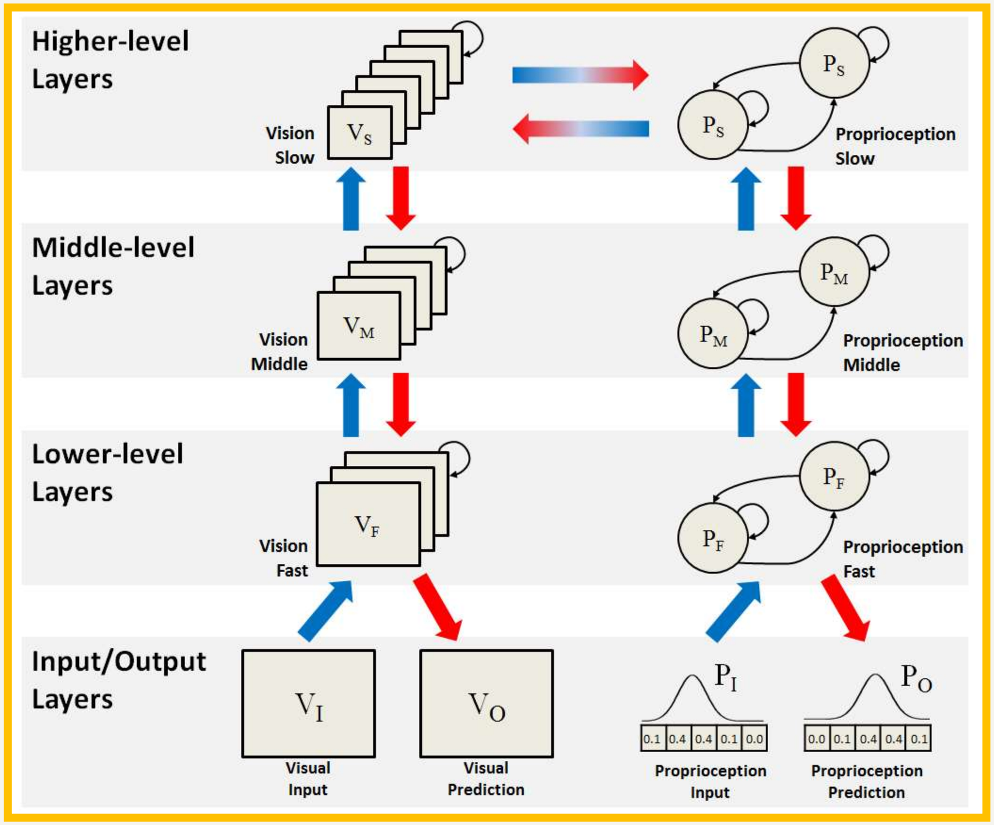
[1]Korea Advanced Institute of Science and Technology
[2]Okinawa Institute of Science and Technology

Overview

"Deep Dynamic Neural Network Model" which can

1. Build a **Predictive Internal Model** of the world from **sensorimotor experience**
   - Predicting dynamic visuo-proprioceptive patterns
2. **Minimize Prediction Error** through **updating internal states** of the neurons
   - Inferring intention of the perceived patterns
   - Recalling visuo-proprioceptive representations

Proposed Model



P-VMDNN

Predictive

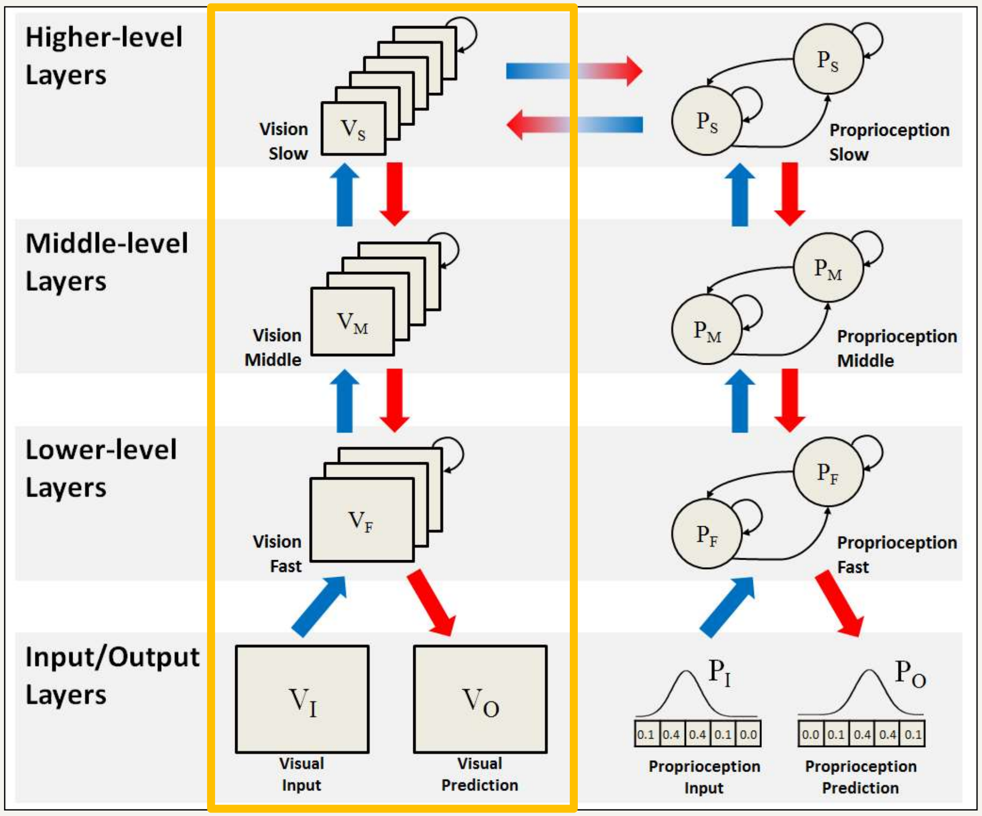
Visuo-**M**otor

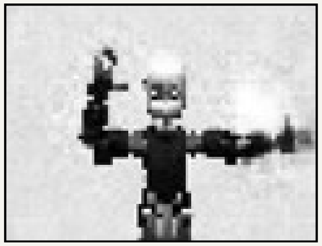
Deep

Dynamic

Neural **N**etwork

Predictive Visuo-Motor Deep Dynamic Neural Network

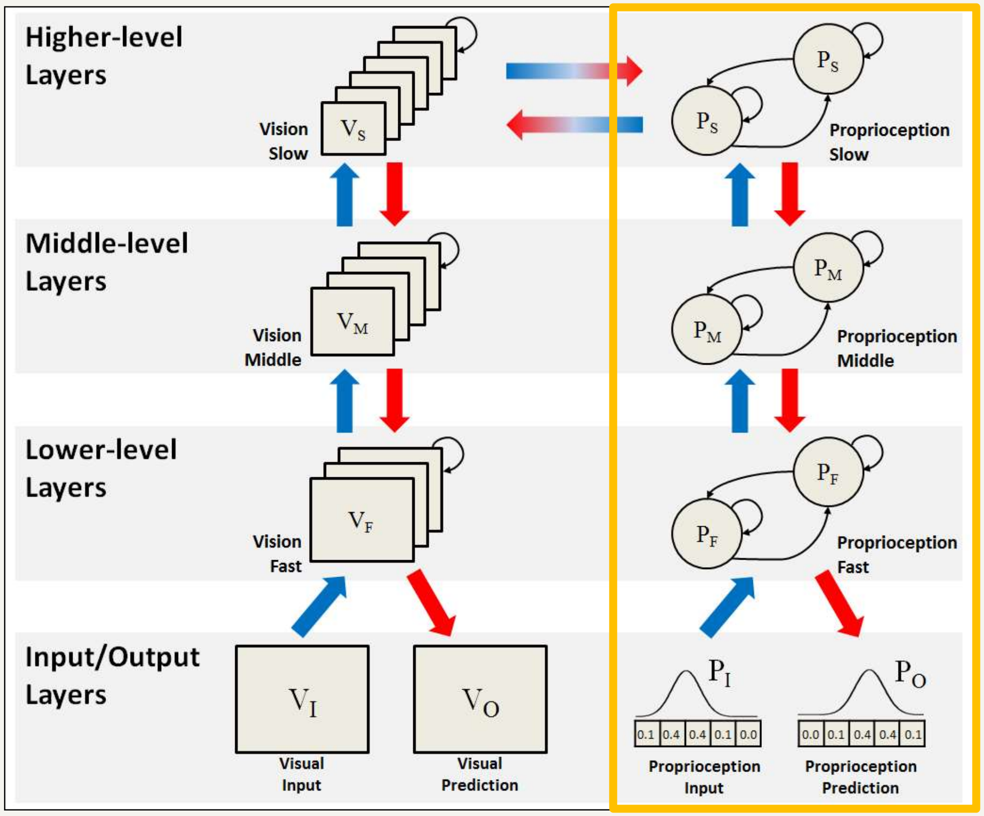


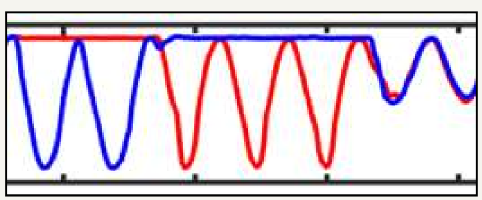


Visual Pathway

- Predicts pixel-level dynamic visual images
  - Implemented by P-MSTRNN (Predictive-Multiple Spatio-Temporal Scales RNN)
  - 4 Layers
    - Vision Input/Output
    - Vision Fast
    - Vision Middle
    - Vision Slow

Predictive Visuo-Motor Deep Dynamic Neural Network

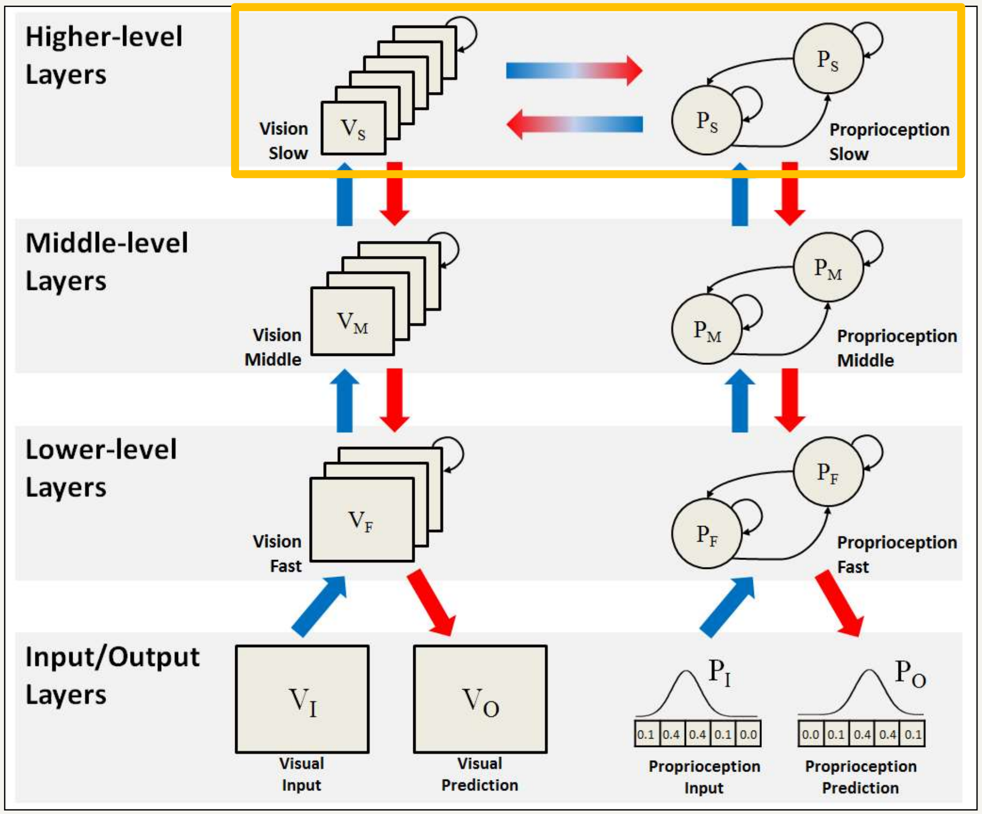




Proprioceptive Pathway

- Predicts robot's joint position values
  - Implemented by MTRNN (Multiple Timescales RNN)
  - 4 Layers
    - Proprioception Input/Output
    - Proprioception Fast
    - Proprioception Middle
    - Proprioception Slow

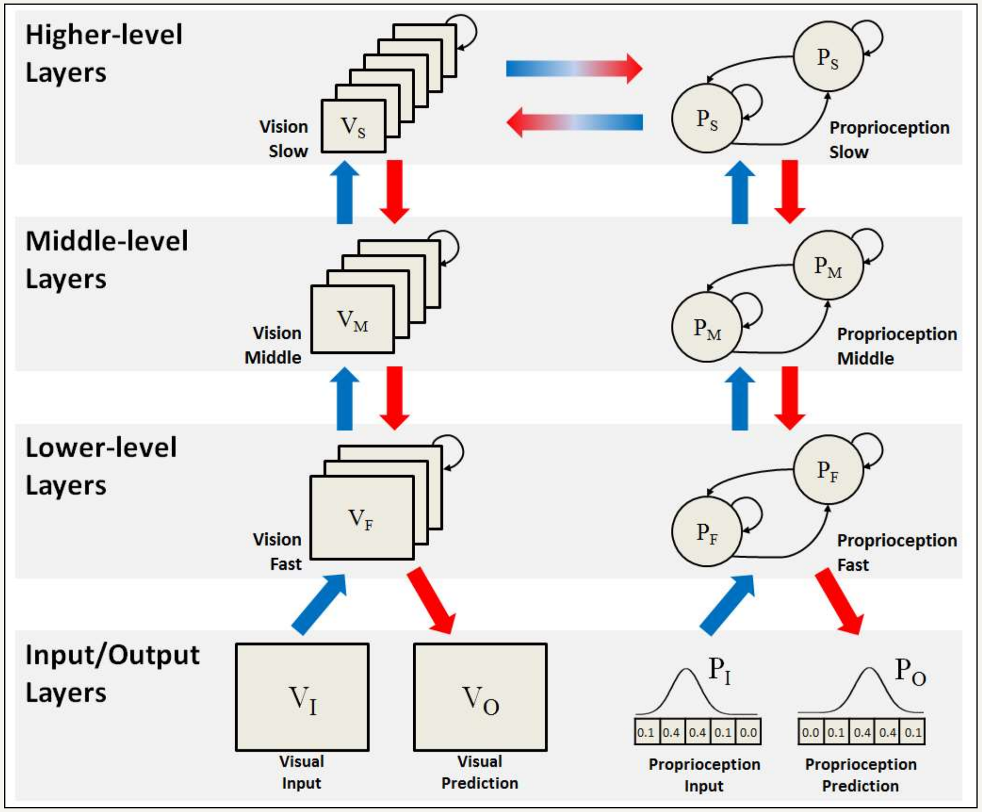
Predictive Visuo-Motor Deep Dynamic Neural Network



Lateral Connection

- Between the highest level of each pathway
- Coupling of Vision & Proprioception
- Trained in a holistic manner
  - End-to-End Training

Predictive Visuo-Motor Deep Dynamic Neural Network



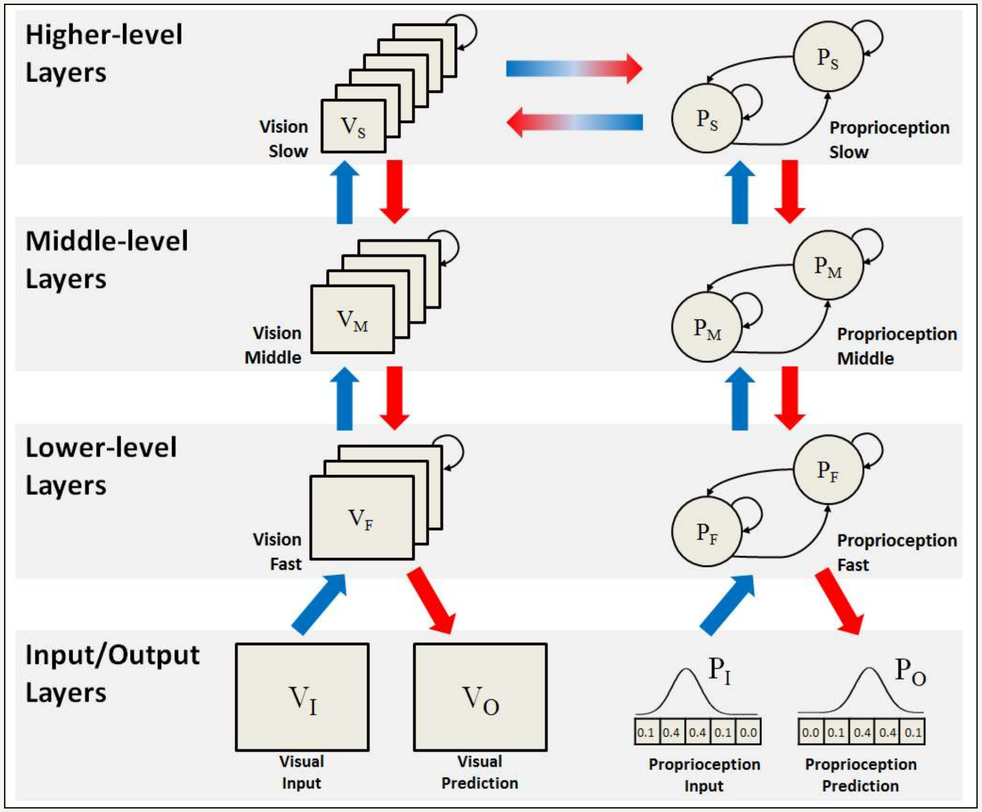
Spatio-Temporal Hierarchy

Larger Time Constants,
Longer Distance Connectivity

Smaller Time Constants,
Shorter Distance Connectivity

"Emergence of
Functional Hierarchy"

Predictive Visuo-Motor Deep Dynamic Neural Network



Key Features

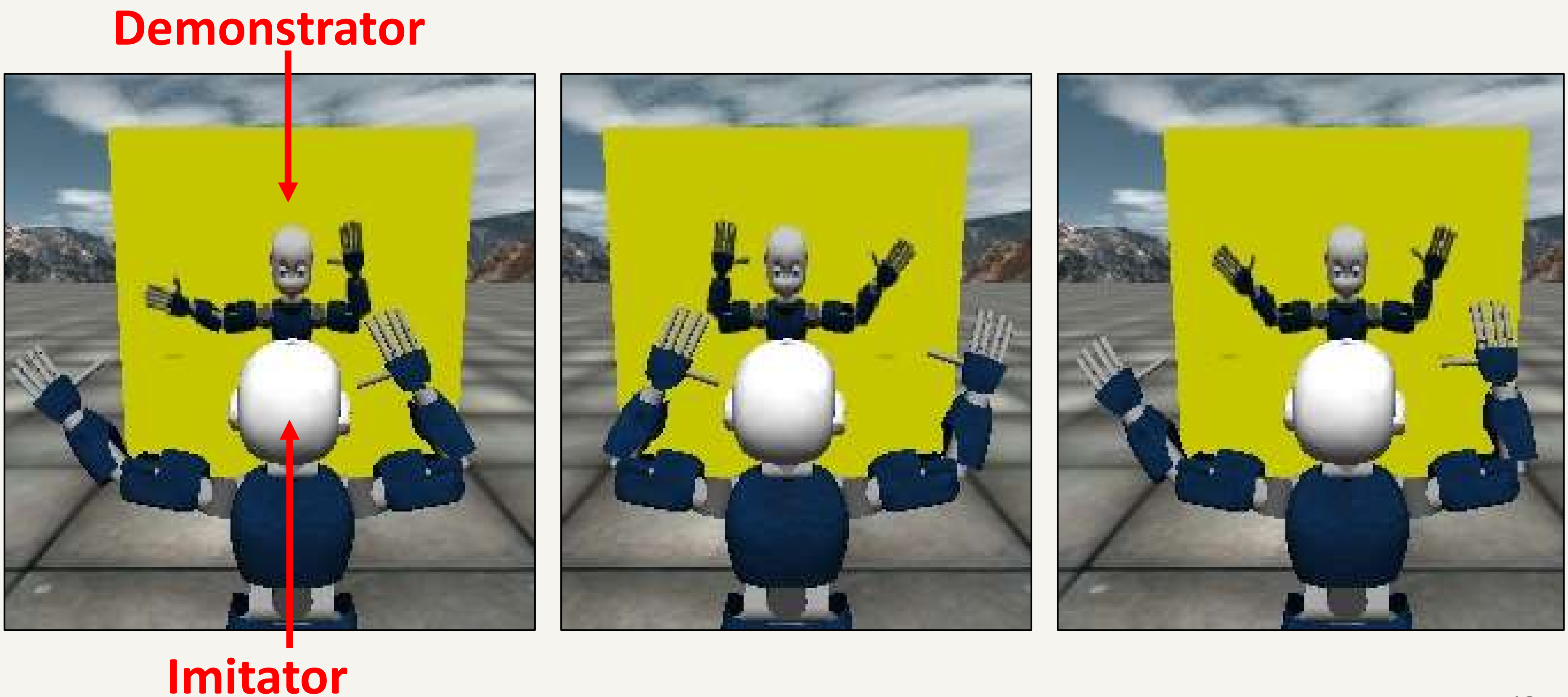
- **Mental Simulation**
- **Prediction Error Minimization**
- Processing of Spatio-Temporal Patterns*
- Coupling of Vision & Proprioception *

(*Hwang et al., ICDL-EPIROB 2016)

- Overview
- Proposed Model
- **Experiment Setting**
- Mental Simulation
- Prediction Error Minimization
- Conclusion

Experiment **Setting**

- Task: **Imitating** another robot’s gestures
  - Robotic Platform: iCub Simulator
  - Visuomotor Coordination, Observing the movements, Understanding the intention, Predicting the next movements



Experiment **Setting**

- Dataset: Acquired From “tutoring” (kinesthetic teaching)
  - 16 hand-waving gestures:
    - Visual Images (64 x 48 grayscale)
    - Joint Position Values (left & right elbows)
- Training: 40,000 epochs, BPTT, ADAM on Tensorflow
- After training:
  - **Connection weights & biases**: <u>same</u> for all training data
  - **Initial states**: <u>different</u> for each training data
- Network Settings:

	Layer	Time Constants	Feature Maps		Top-Down Kernel		Bottom-Up Kernel		Recurrent Kernel		Lateral Kernel	
			Number	Size	Size	Stride	Size	Stride	Size	Stride	Size	Stride
Visual Pathway	V_F	2	4	60×44	4×4	2,2	5×5	1,1	2×2	1,1	-	-
	V_M	4	8	29×21	5×5	2,2	4×4	2,2	2×2	1,1	-	-
	V_S	8	12	13×9	-	-	5×5	2,2	2×2	1,1	13×9	1,1

	Layer	Time Constants	Number of Neurons	Top-Down Weights	Bottom-Up Weights	Recurrent Weights	Lateral Kernel	
Proprio-ceptive pathway	P_F	2	30	30×20	30×20	30×30	-	-
	P_M	4	20	20×10	20×30	20×20	-	-
	P_S	8	10	-	10×20	10×20	13×9	1,1

Experiment Structure

- **Exp 1. Mental Simulation**
  - Imagining possible outcome of action
  - Without external inputs, but with given intention
- **Exp 2. Prediction Error Minimization**
  - Exp 2-1. Minimizing **Visual** Prediction Error
  - Exp 2-2. Minimizing **Proprioceptive** Prediction Error

Mental Simulation of Action

Mental Simulation

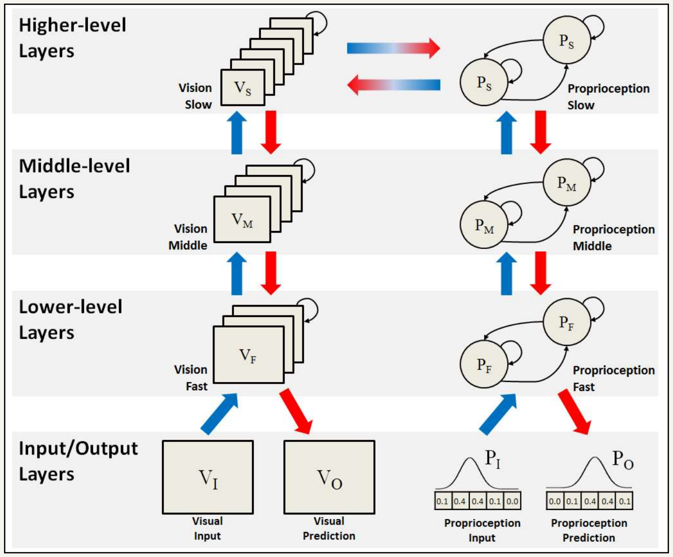
- Ability to imagine probable result of our actions
- Important in social interaction
- Need to provide "a goal" – what to simulate

Implementation in Our Model

- Proactively generating visuo-proprioceptive patterns
  - Without external inputs, but with given intention states
- Anticipating
  - its own action (i.e. Proprioceptive Prediction – joint position values)
  - & others' action (i.e. Visual Prediction – gray scale images)

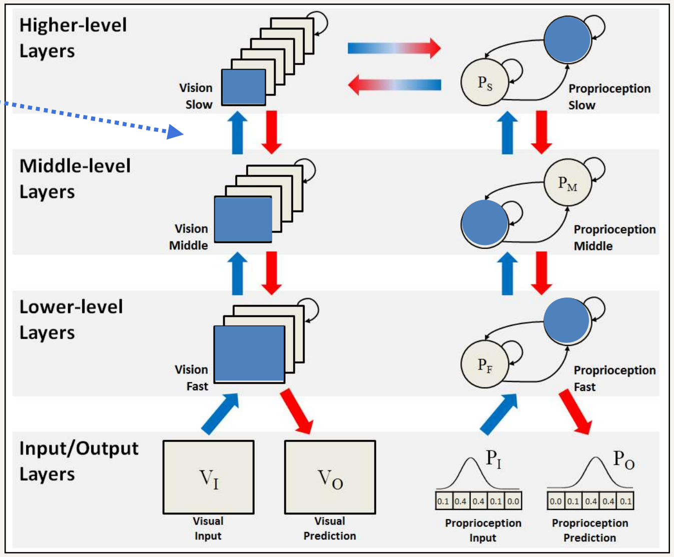
Mental Simulation of Action

- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. Iterate (2) – (3)



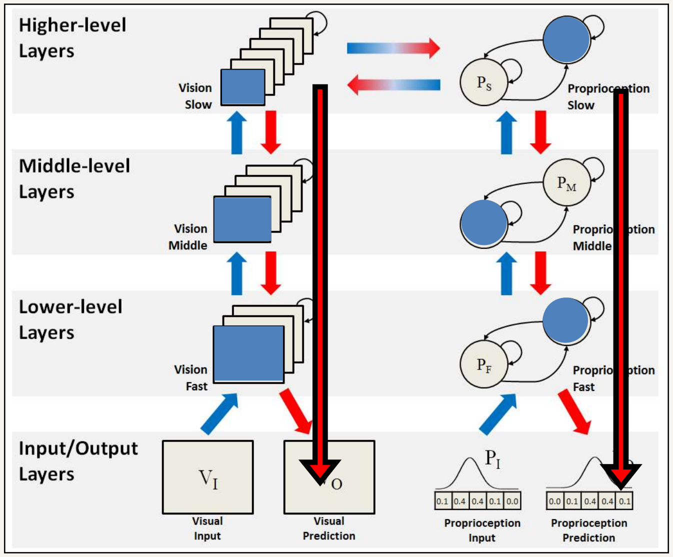
Mental Simulation of Action

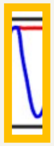
- **Mental Simulation in the Proposed Model**
  1. **Set the "Intention"**
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. Iterate (2) – (3)



Mental Simulation of Action

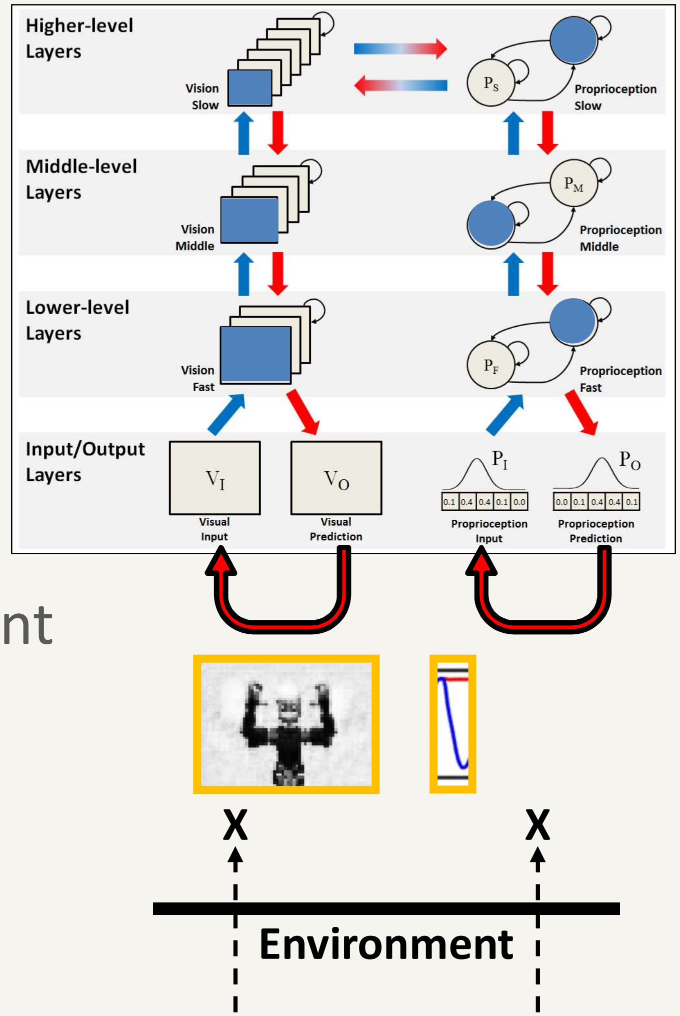
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. **Generate Output**
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. Iterate (2) – (3)





Mental Simulation of Action

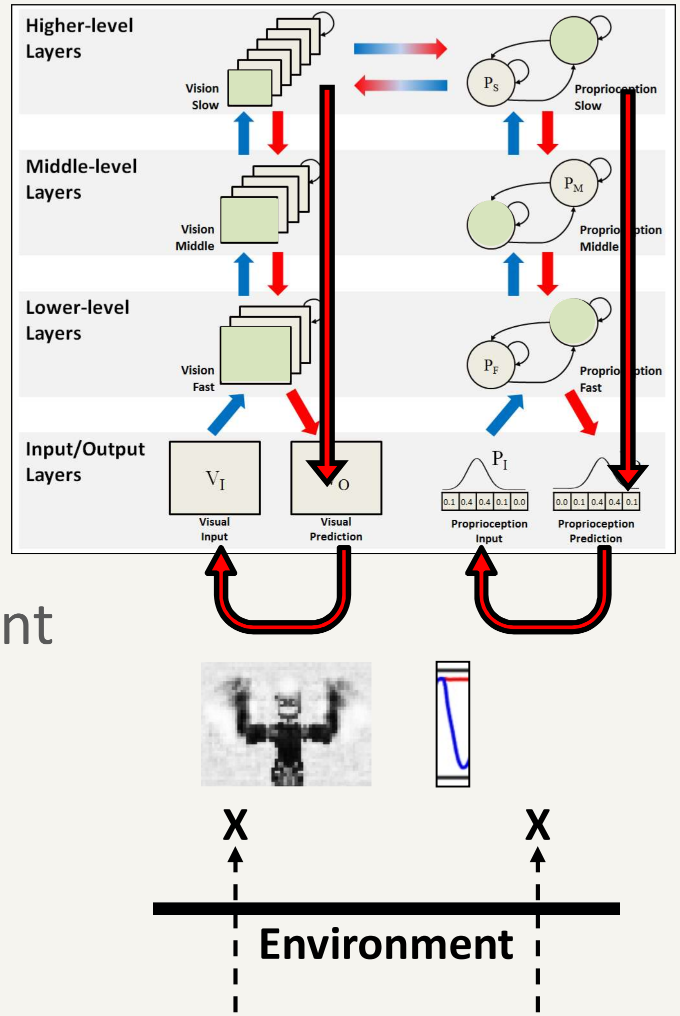
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. **Feed Prediction Output into Input**
     - "Closed-loop Generation"
  4. Iterate (2) – (3)



- Without external input from environment
- Only with given intention

Mental Simulation of Action

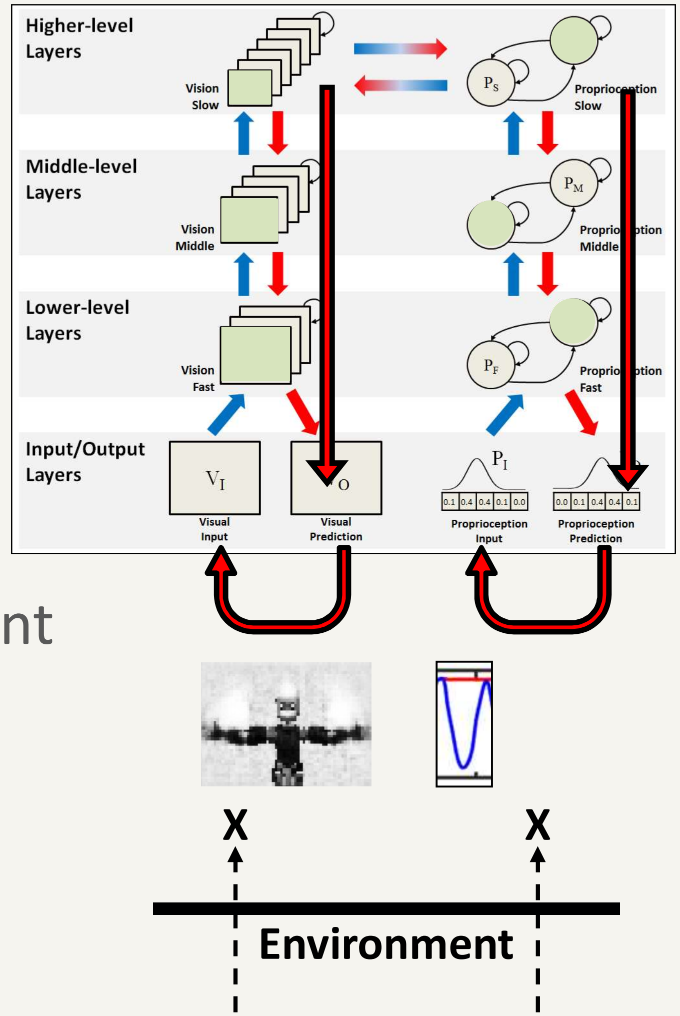
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. **Iterate (2) – (3)**



- Without external input from environment
- Only with given intention

Mental Simulation of Action

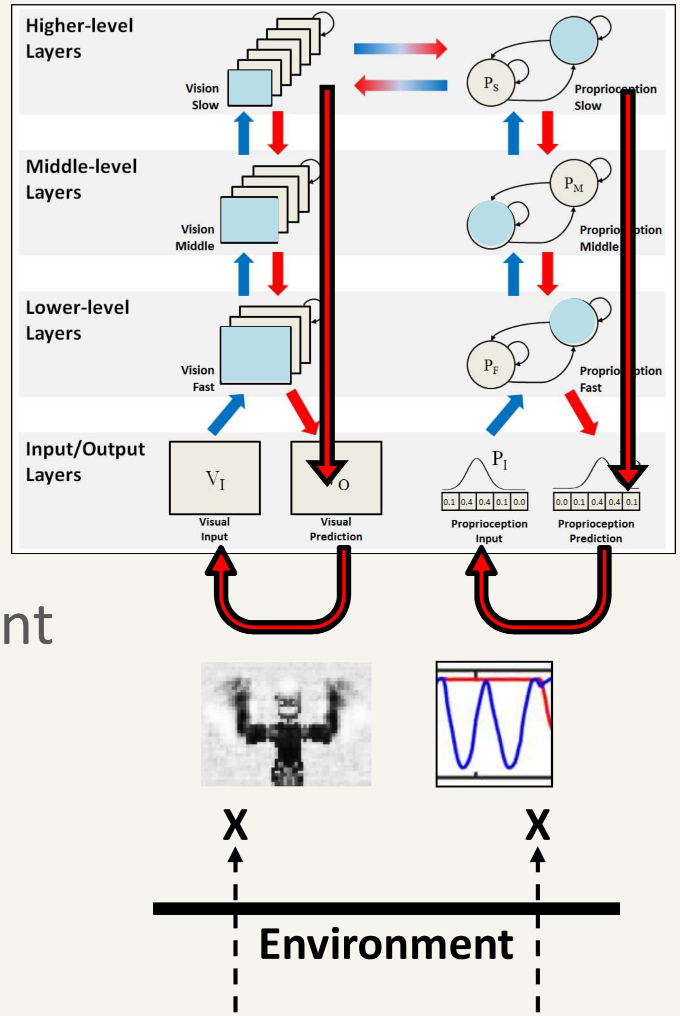
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. **Iterate (2) – (3)**



- Without external input from environment
- Only with given intention

Mental Simulation of Action

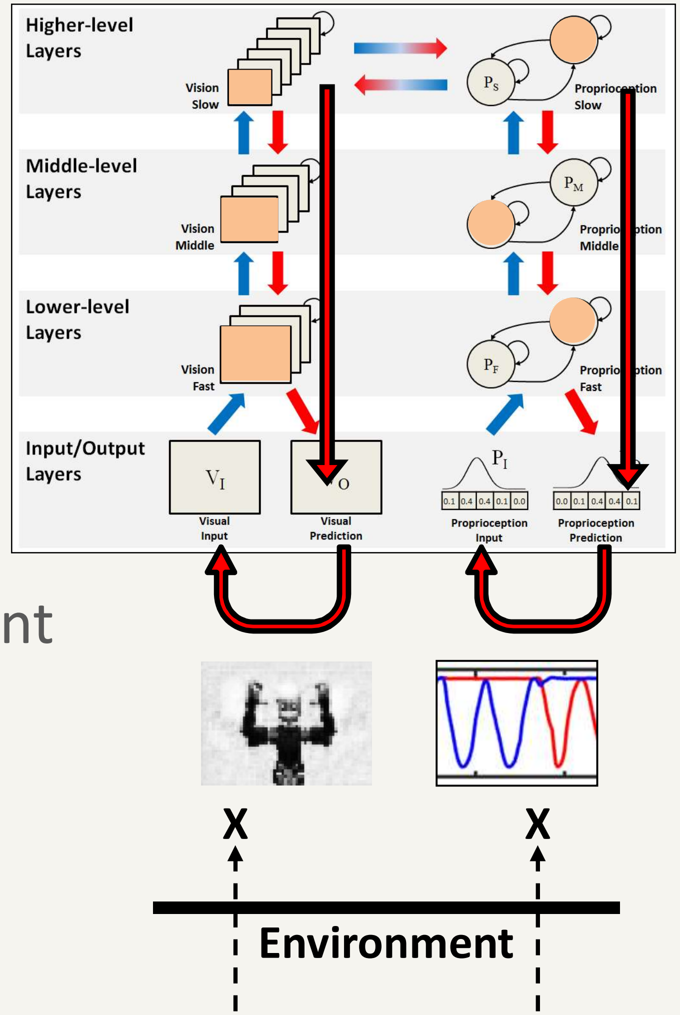
- **Mental Simulation in the Proposed Model**
  1. Set the "Intention"
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - "Closed-loop Generation"
  4. **Iterate (2) – (3)**



- Without external input from environment
- Only with given intention

Mental Simulation of Action

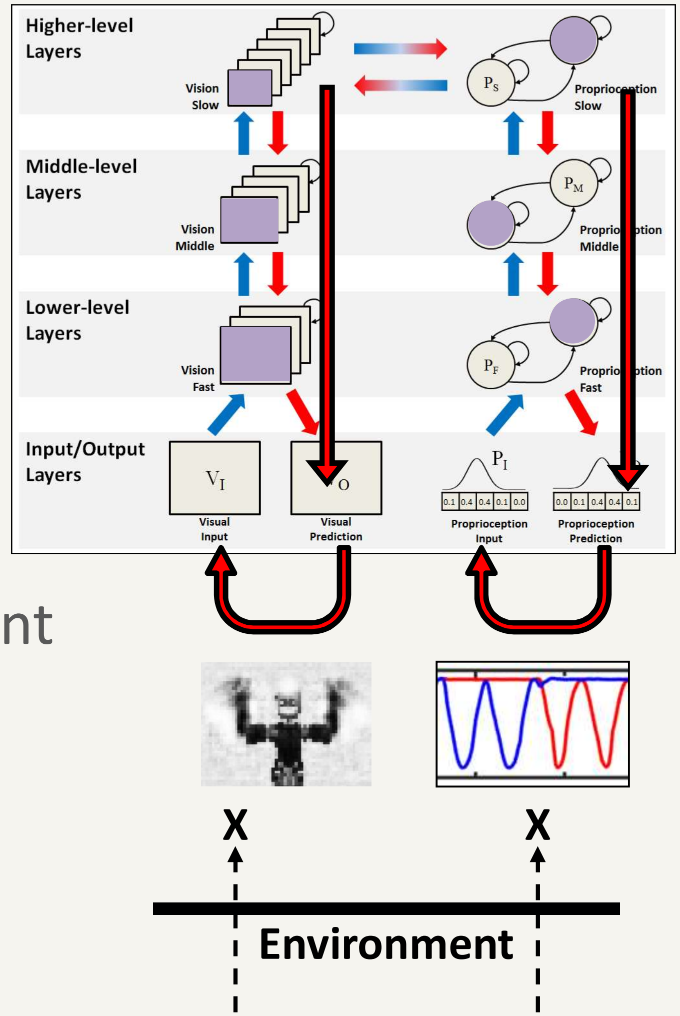
- **Mental Simulation in the Proposed Model**
    1. Set the "Intention"
        - Specified as the initial states
    2. Generate Output
        - Visual & Proprioceptive predictions
    3. Feed Prediction Output into Input
        - "Closed-loop Generation"
    4. **Iterate (2) – (3)**



- Without external input from environment
- Only with given intention

Mental Simulation of Action

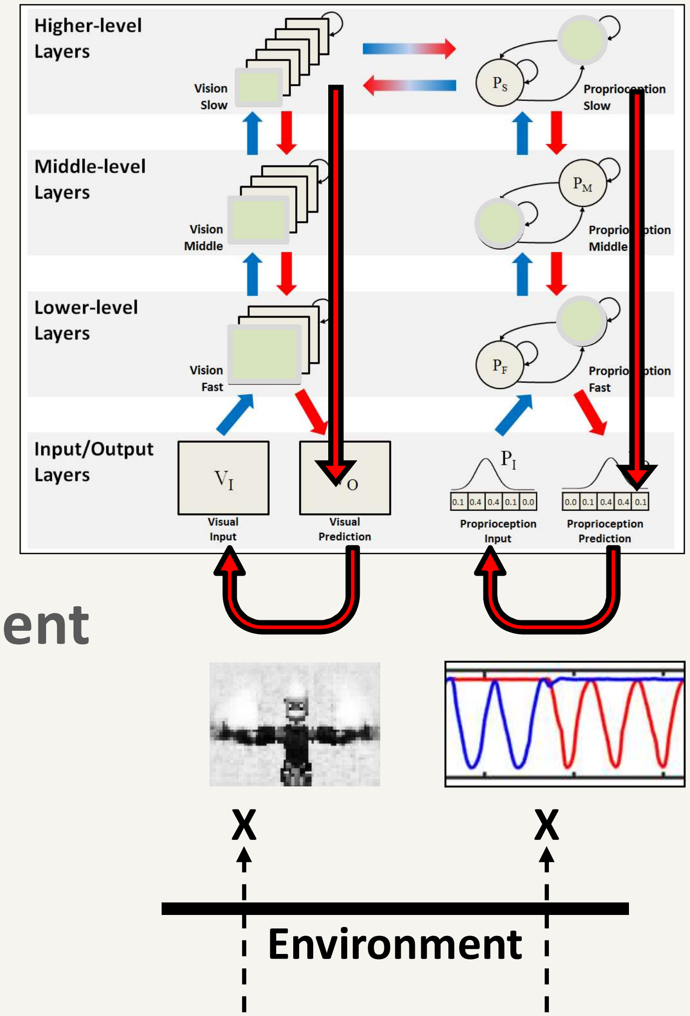
- **Mental Simulation in the Proposed Model**
  1. Set the “Intention”
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - “Closed-loop Generation”
  4. **Iterate (2) – (3)**



- Without external input from environment
- Only with given intention

Mental Simulation of Action

- **Mental Simulation in the Proposed Model**
  1. Set the “Intention”
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - “Closed-loop Generation”
  4. **Iterate (2) – (3)**

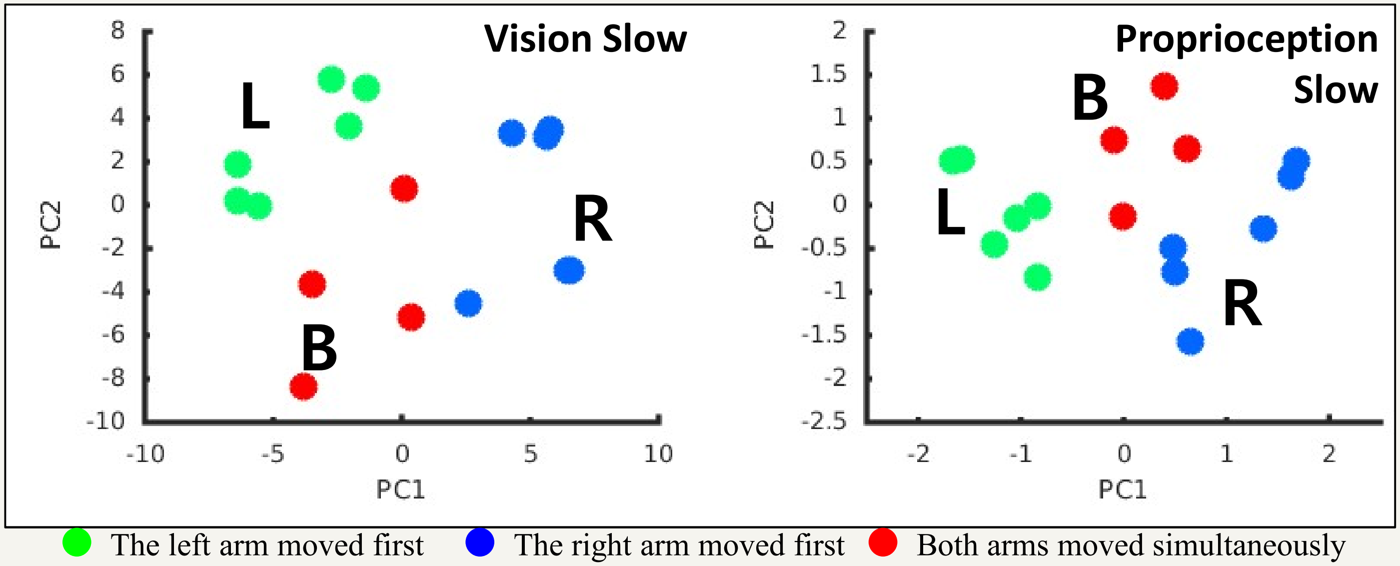


- **Without external input from environment**
- **Only with given intention**

Result

Initial States obtained from Training

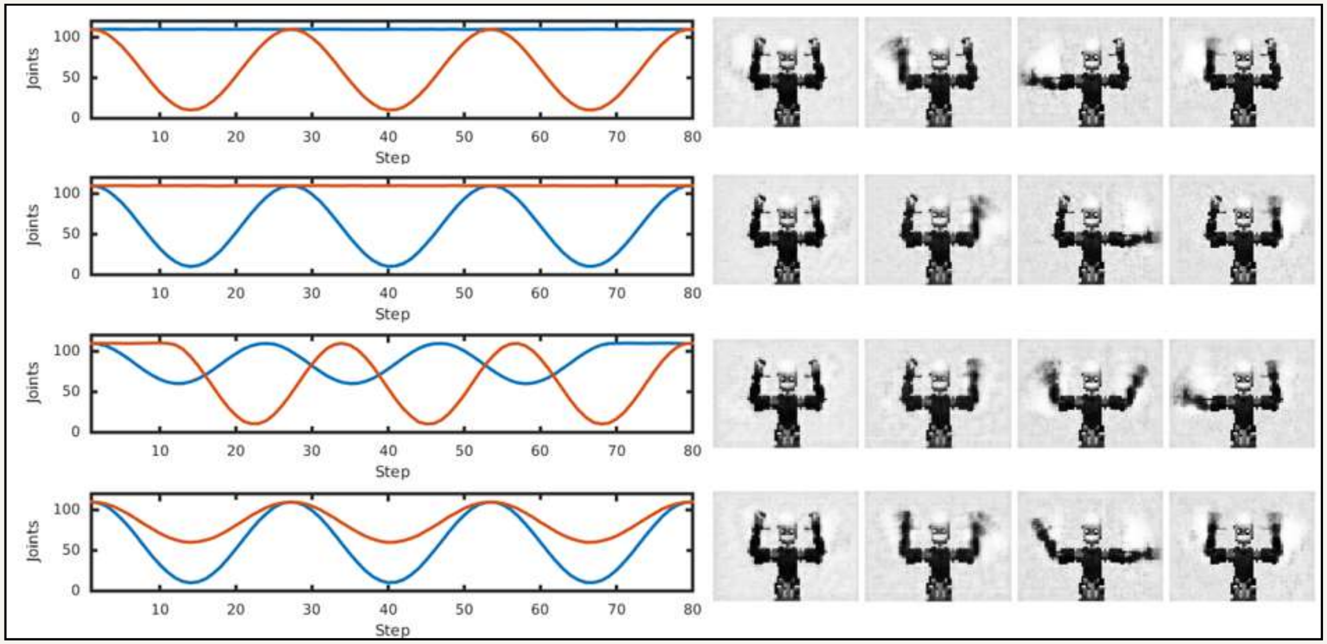
- Different initial states for each training data
- Self-organized higher-level initial states
  - Reflecting the characteristics of the gestures



Result

Mental Simulation of Action

- With given ‘intention’, the model generated visuo-proprioceptive patterns without external inputs
  - Coherent visual and proprioceptive predictions
    - ➔ Vision and Proprioception were tightly coupled



Closed-loop generation of patterns

Overview

Proposed Model

Experiment Setting

Mental Simulation

Prediction Error Minimization

Conclusion

Prediction Error (PE) Minimization

Prediction Error Minimization

- Core of “Predictive Coding”
- Account for MNS (Mirror Neuron System)
  - Recognizing intention from observation by minimizing prediction error at the levels of a cortical hierarchy (Kilner et al., 2007)

Implementation in Our Model

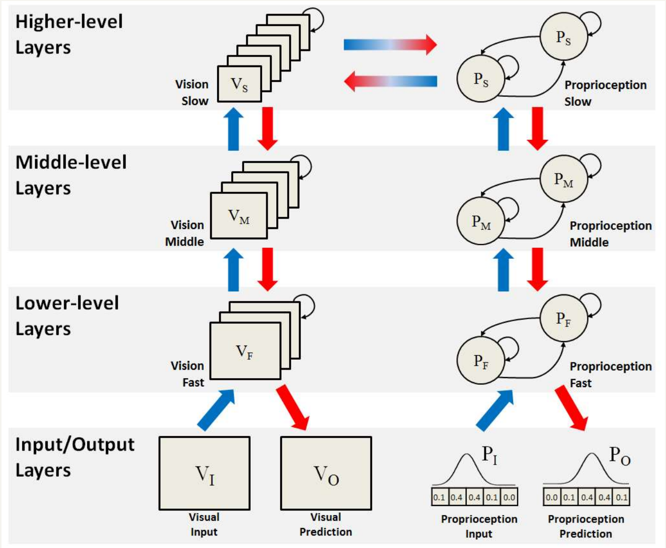
- Recognizing intention of the perceived patterns by minimizing prediction error

  <HOW>➔ Updating internal states of neurons at each level of the hierarchy

“Error Regression Scheme”

Prediction Error (PE) Minimization

- **Error Regression Scheme (ERS)**
  - Implementation of PE Minimization
    (Tani, 2016)



At each time step

1. Generate Visuo-Prop. Predictions
   - Top-Down Process
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times

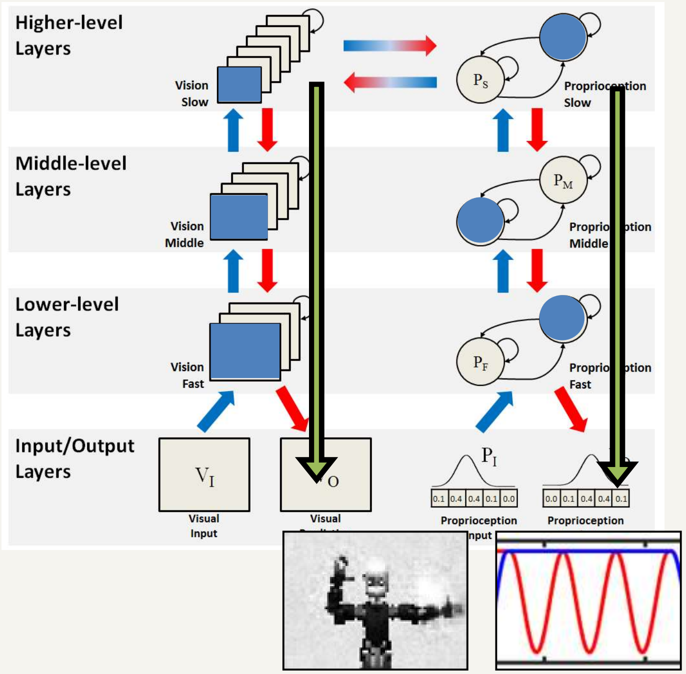
Environment

Prediction Error (PE) Minimization

- **Error Regression Scheme (ERS)**
  - Implementation of PE Minimization
    (Tani, 2016)

At each time step

1. **Generate Visuo-Prop. Predictions**
   - Top-Down Process
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times



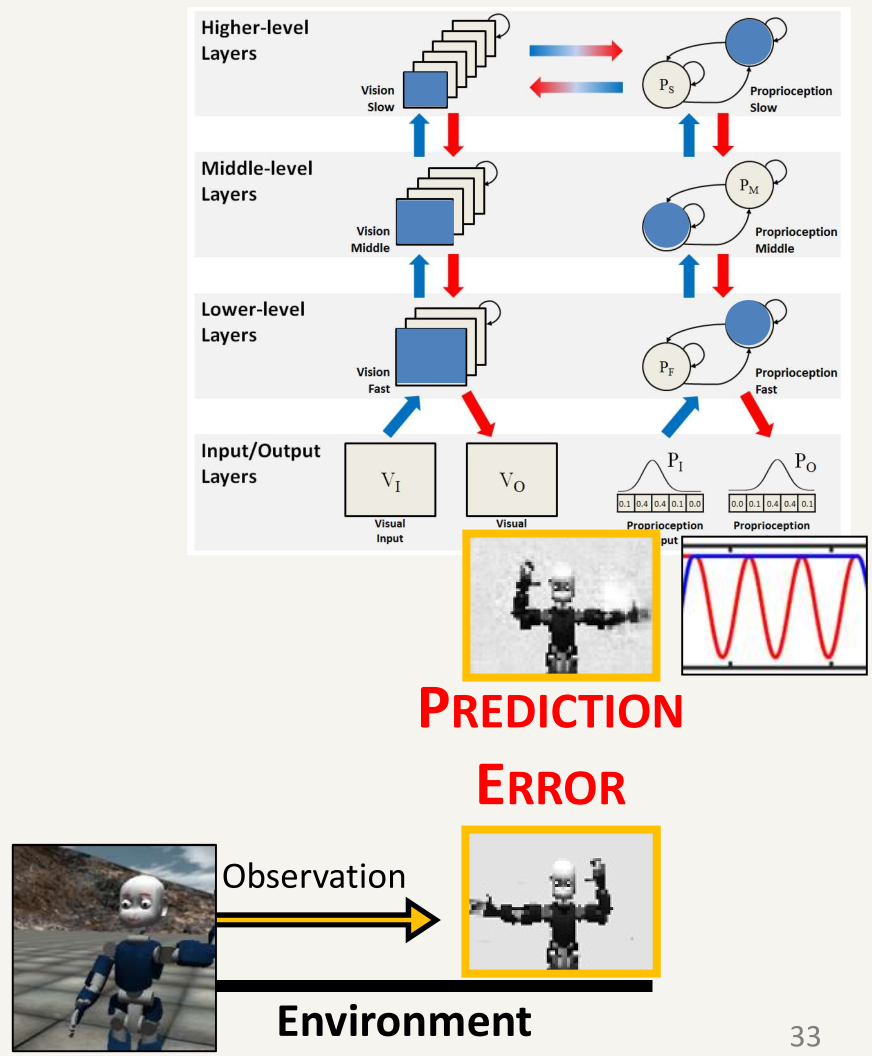
Environment

Prediction Error (PE) Minimization

- **Error Regression Scheme (ERS)**
  - Implementation of PE Minimization
    (Tani, 2016)

At each time step

1. Generate Visuo-Prop. Predictions
   - Top-Down Process
2. **Compute Prediction Error**
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times

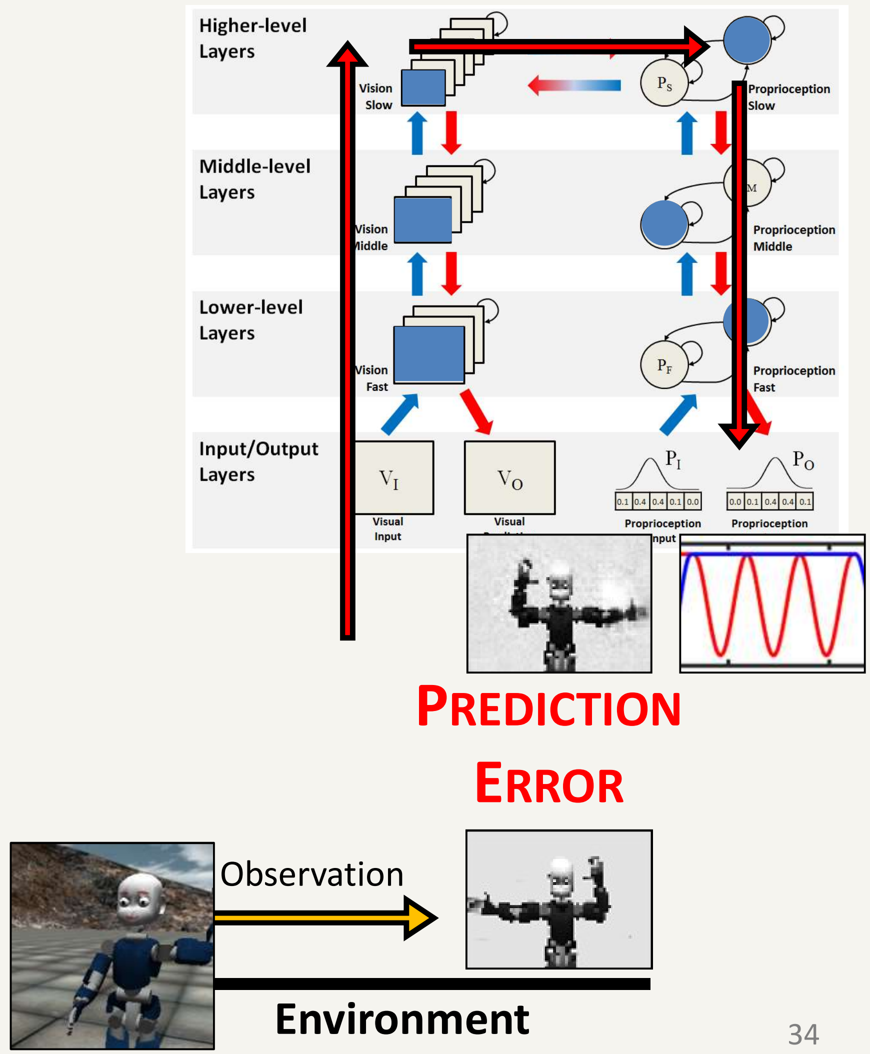


Prediction Error (PE) Minimization

- **Error Regression Scheme (ERS)**
  - Implementation of PE Minimization
    (Tani, 2016)

At each time step

1. Generate Visuo-Prop. Predictions
   - Top-Down Process
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. **Backpropagate Prediction Error** & Update Intention State
   - Bottom-Up Process
4. Iterates a Few Times

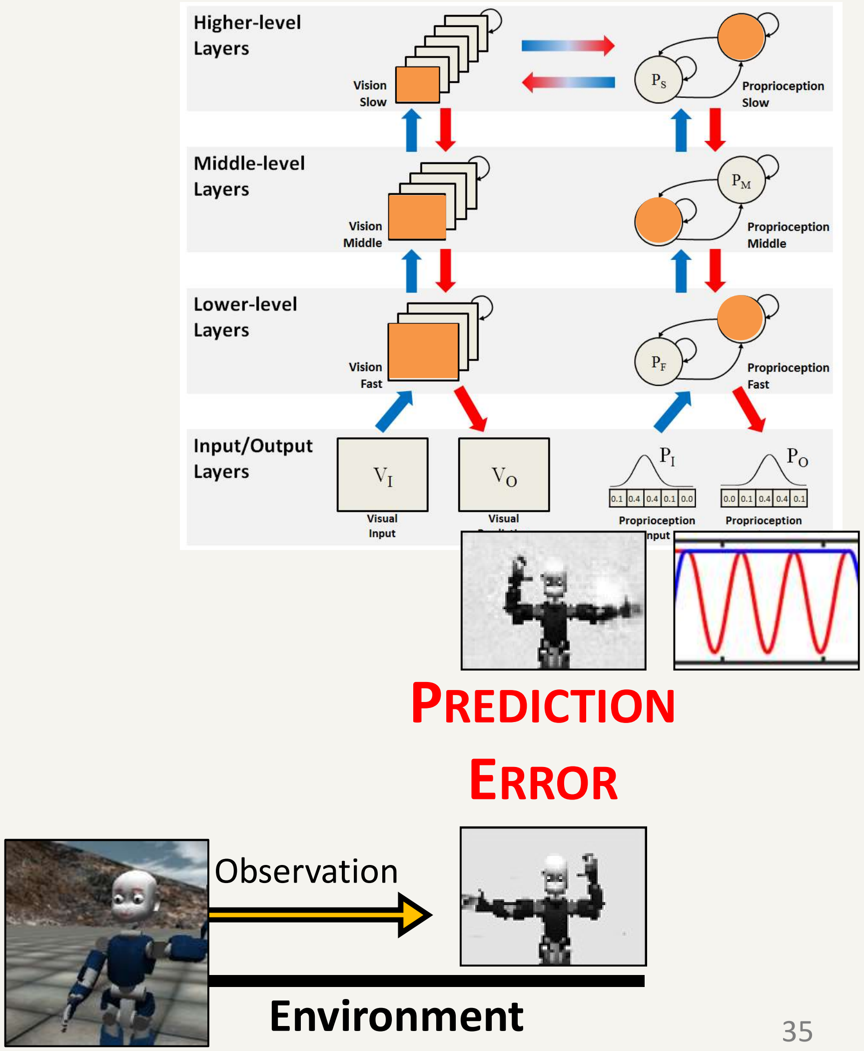


Prediction Error (PE) Minimization

- **Error Regression Scheme (ERS)**
  – Implementation of PE Minimization
  (Tani, 2016)

At each time step

1. Generate Visuo-Prop. Predictions
   - Top-Down Process
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & **Update Intention State**
   - Bottom-Up Process
4. Iterates a Few Times

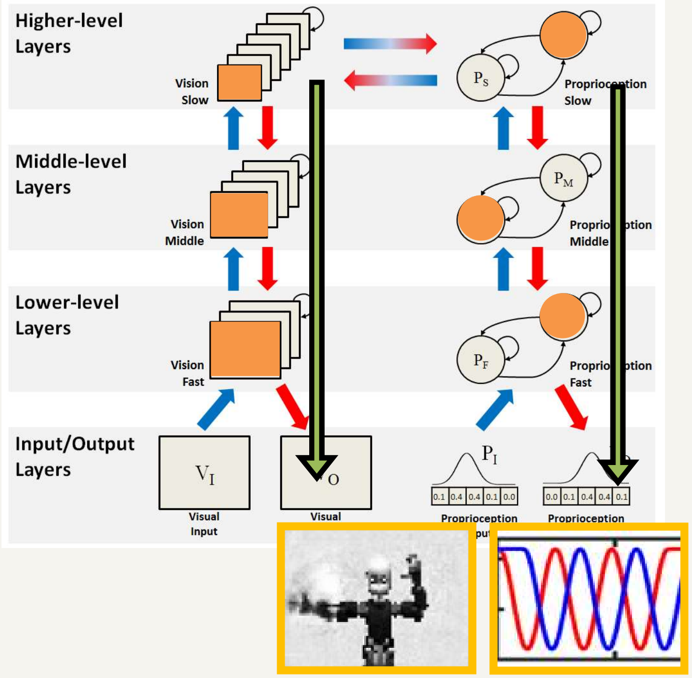


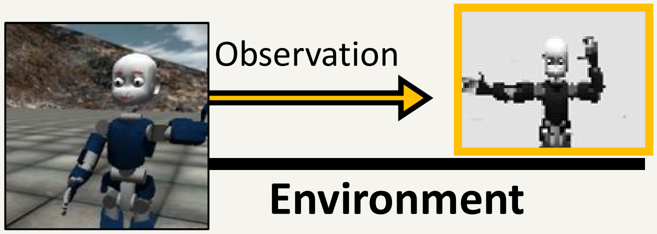
Prediction Error (PE) Minimization

- **Error Regression Scheme (ERS)**
  - Implementation of PE Minimization
    (Tani, 2016)

At each time step

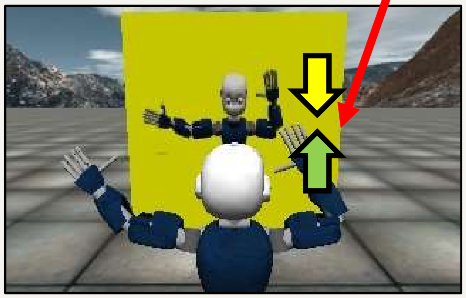
1. Generate Visuo-Prop. Predictions
   - Top-Down Process
2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   - Bottom-Up Process
4. **Iterates a Few Times**





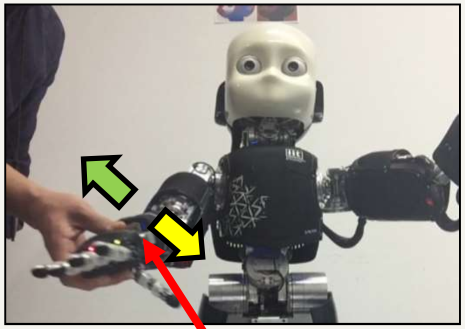
Experimental **Conditions**

VISUAL PREDICTION ERROR



1. Minimizing “**Visual**” Prediction Error
   – Minimizing the difference between
     - Visual Prediction (i.e., predicted gesture)
     - Observation (i.e., observed gesture)

2. Minimizing “**Proprioceptive**” Prediction Error
   – Minimize the difference between
     - Prop. Prediction (i.e., predicted joint position values)
     - Observation (i.e., perceived position values)



PROP. PREDICTION ERROR

Testing Environment

- A target sequence consisting of 4 concatenated patterns

Result

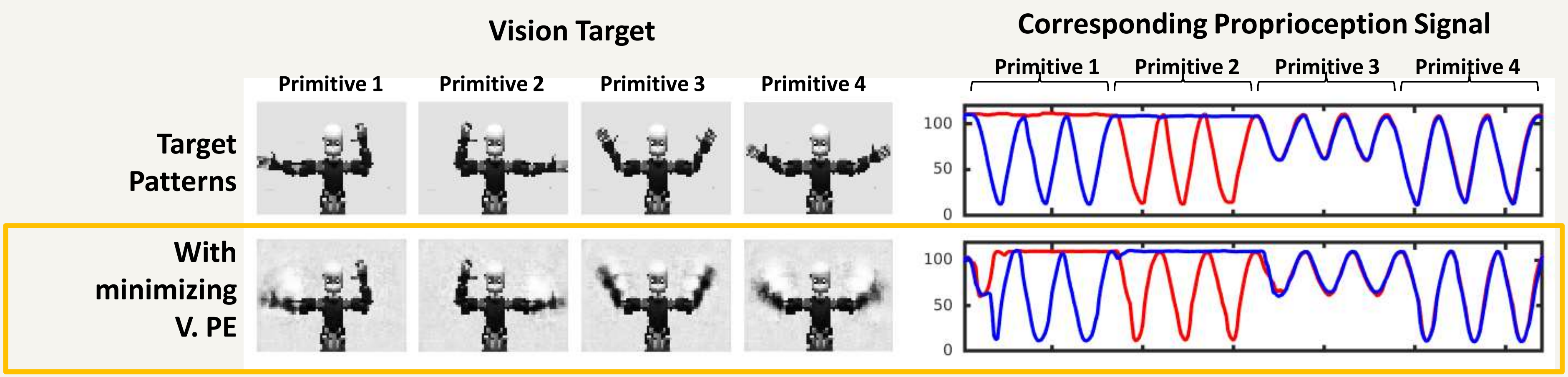
1) Minimizing Visual Prediction Error

- Vision target is given
- The model minimizes Visual PE

- Corresponding Prop. Signal
- (for imitation)

Result
1) Minimizing Visual Prediction Error

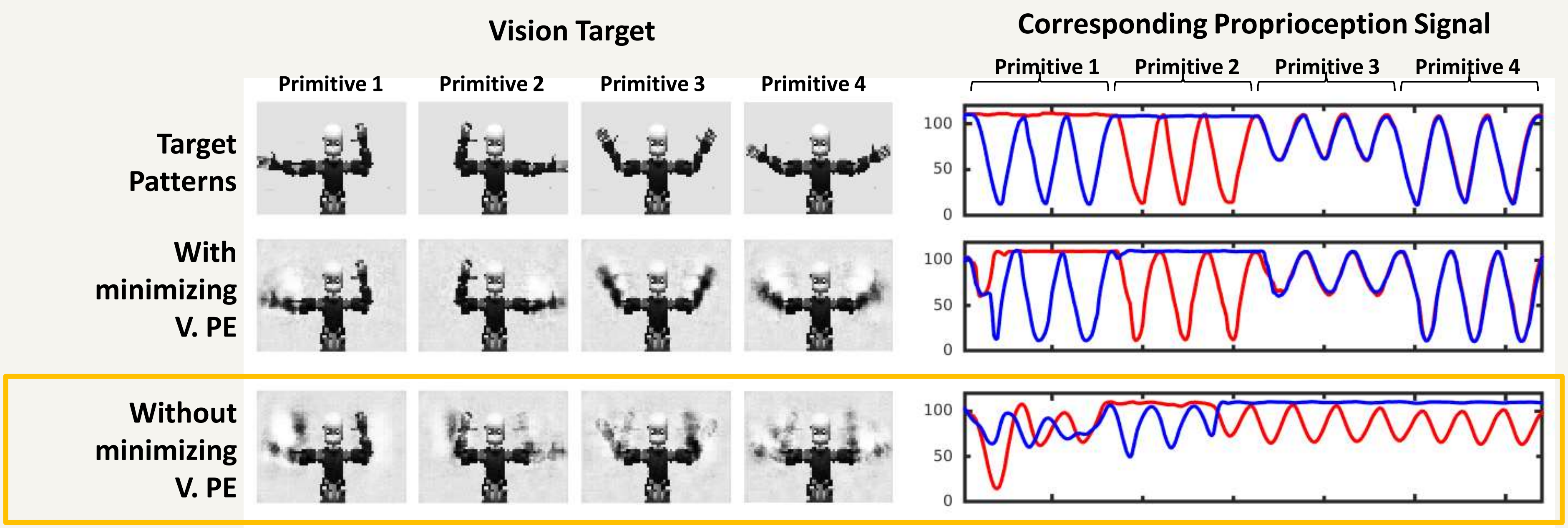
- **With Visual PE Minimization**
  - Successfully predicted visual images
  - Generated corresponding Proprioceptive Prediction ➔ **Successful imitation**
- Without Visual PE Minimization
  - Did NOT predict Visual Images & Proprioceptive Signals ➔ Unsuccessful imitation



Result

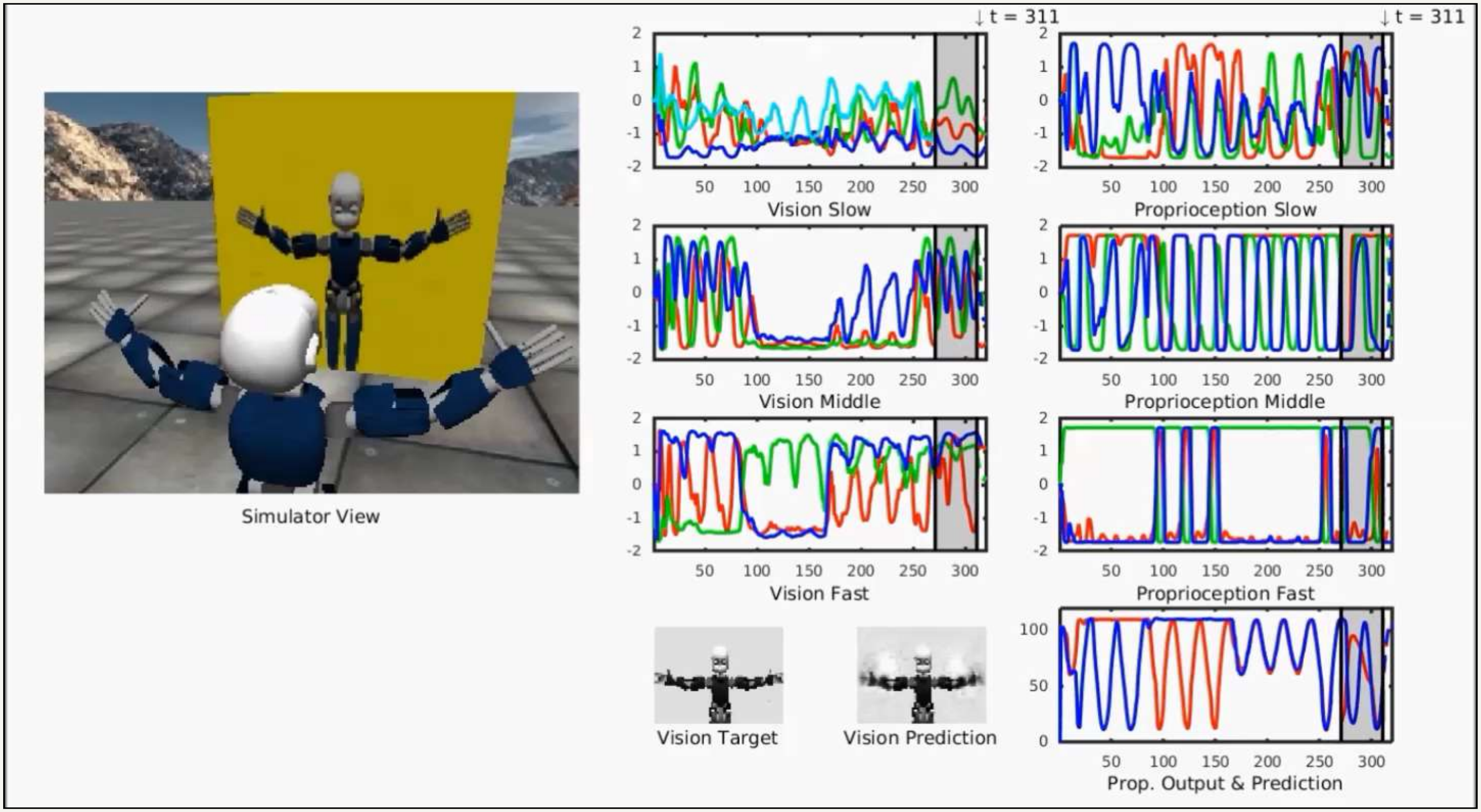
1) Minimizing Visual Prediction Error

- With Visual PE Minimization
  - Successfully predicted visual images
  - Generated corresponding Proprioceptive Prediction ➔ Successful imitation
- **Without Visual PE Minimization**
  - Did NOT predict Visual Images & Proprioceptive Signals ➔ **Unsuccessful imitation**



Result
1) Minimizing Visual Prediction Error

- Video

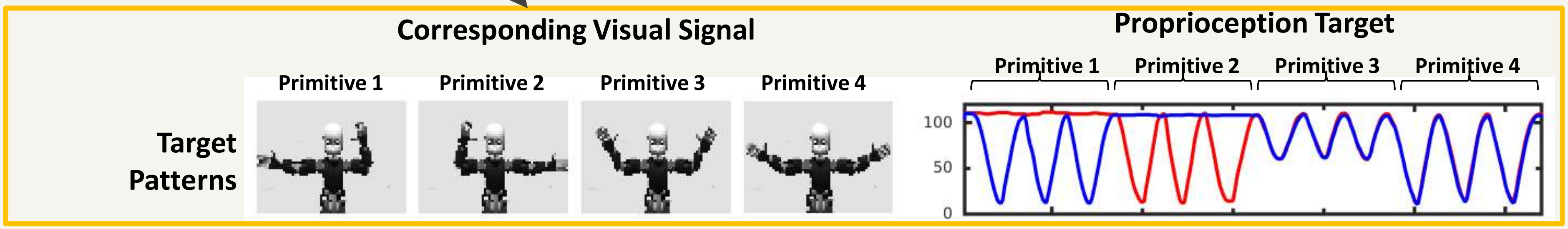


Result

2) Minimizing Proprioceptive Prediction Error

- Corresponding Visual Image

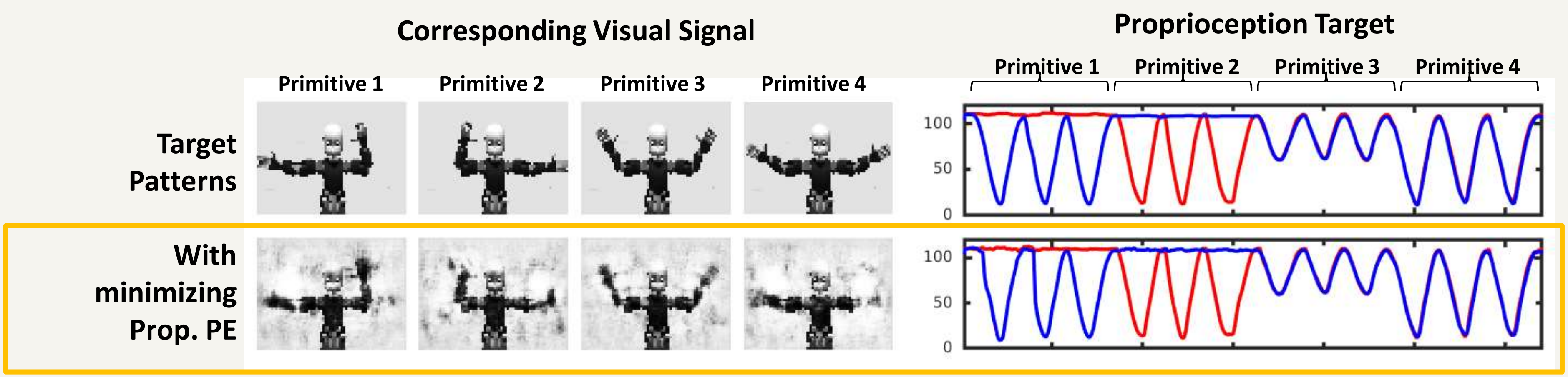
- Prop. target is given
- The model minimize PropPE



Result

2) Minimizing Proprioceptive Prediction Error

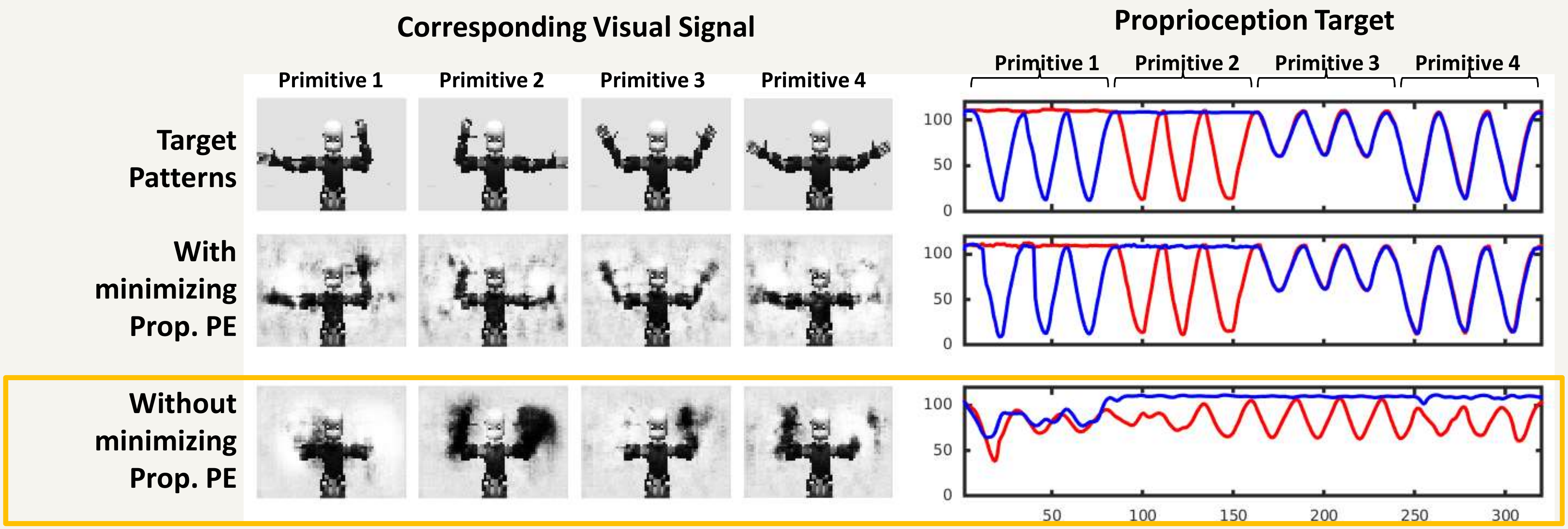
- **With Prop.PE Minimization**
  - Successfully minimized Prop.PE
  - **Generated corresponding Visual Prediction (visual imaginary)**
- Without Prop.PE Minimization
  - Not able to adapt to incoming Prop. Signal / generate corresponding visual imaginary



Result

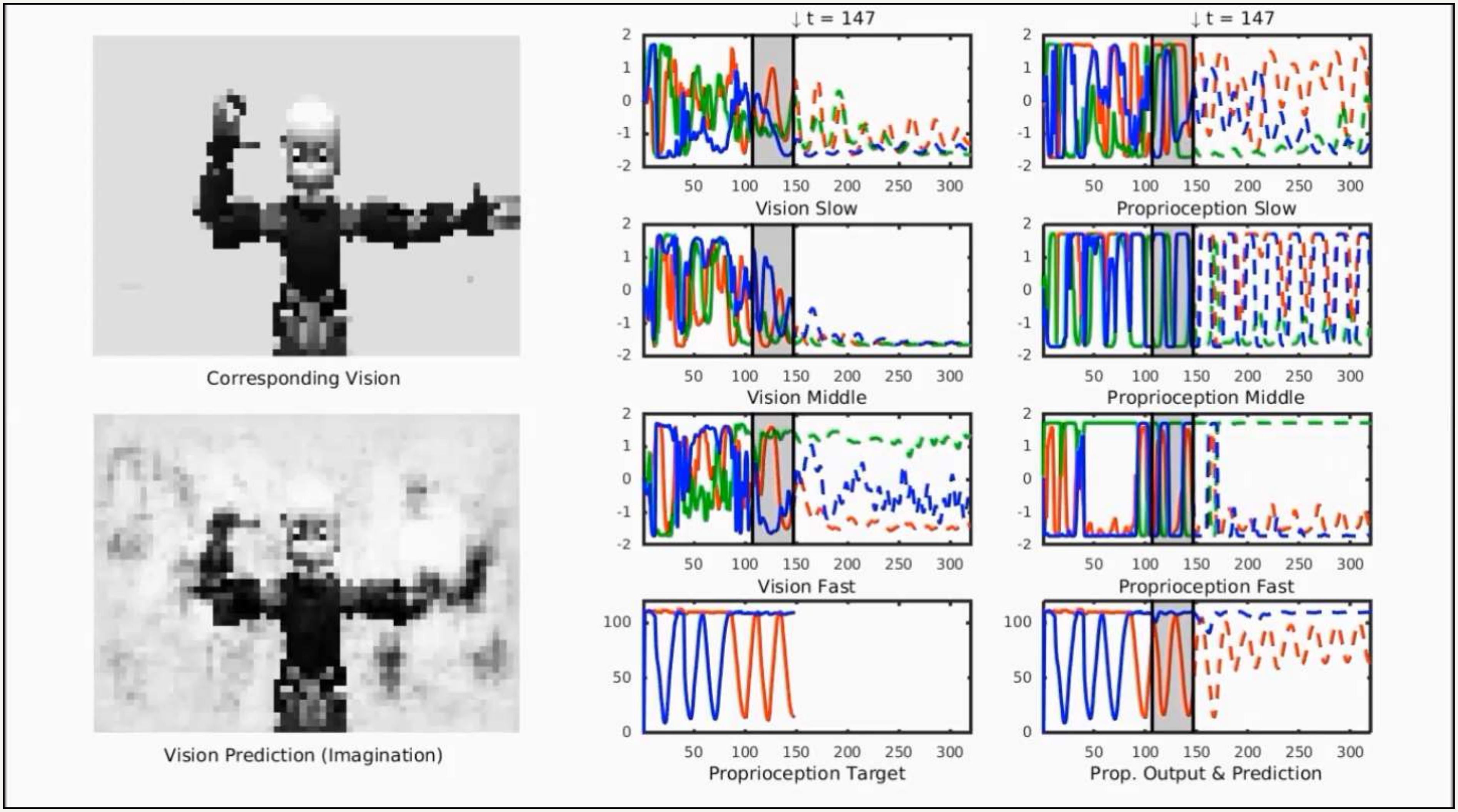
2) Minimizing Proprioceptive Prediction Error

- With Prop.PE Minimization
  - Successfully minimized Prop.PE
  - Generated corresponding Visual Prediction (visual imaginary)
- **Without Prop.PE Minimization**
  - **Not able to adapt to incoming Prop. Signal / generate corresponding visual imaginary**



Result
2) Minimizing **Proprioceptive** Prediction Error

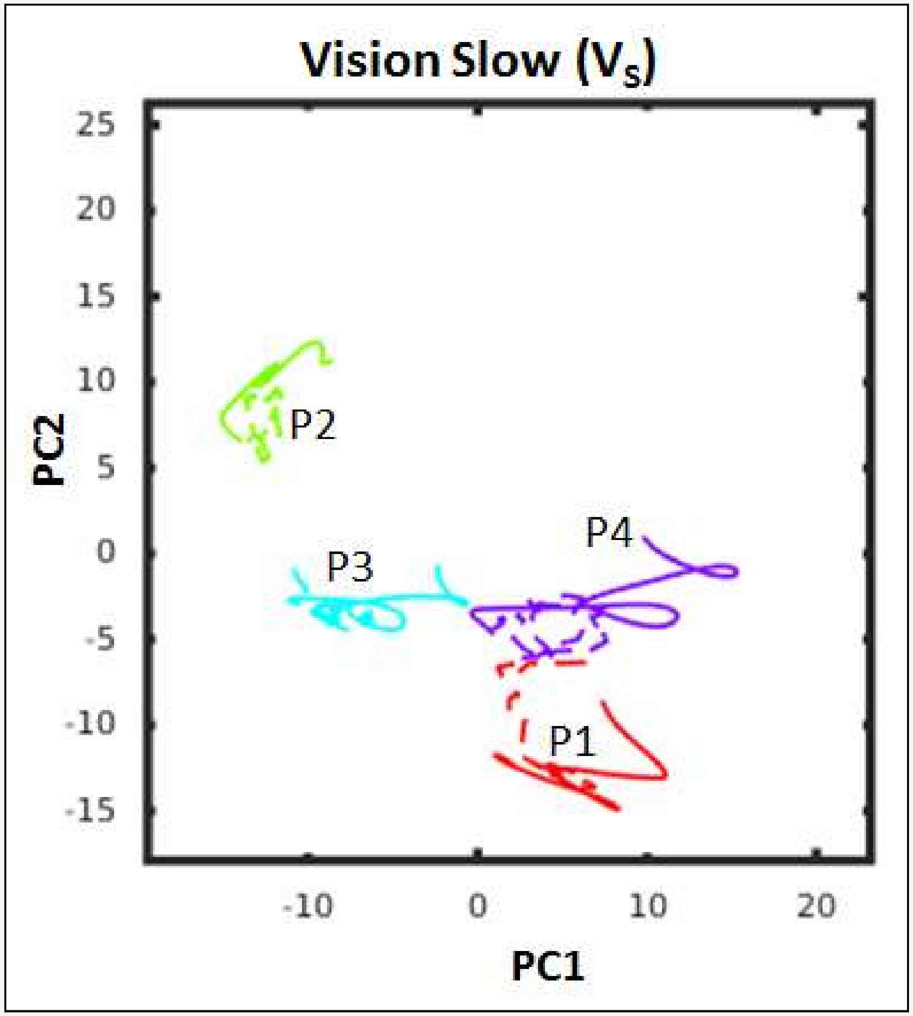
- VIDEO



Result

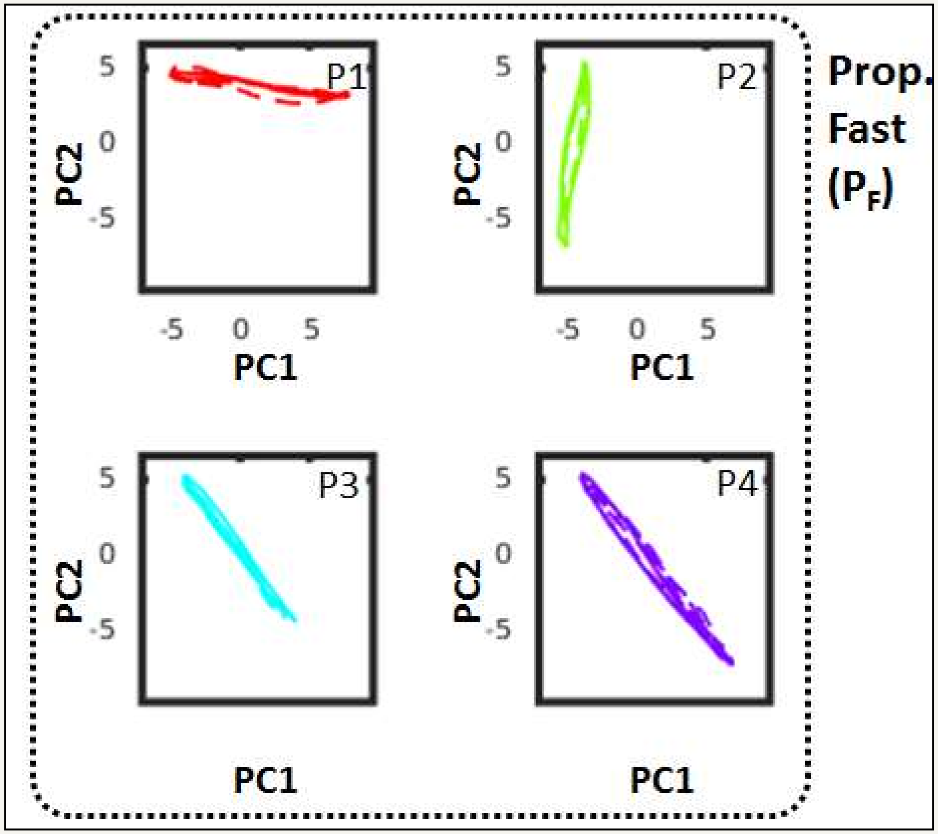
Neural Activation while Minimizing Visual Prediction Error

- **Trajectory of Neural Activation**
  - **Dashed Lines:** During **PE Minimization**
  - **Solid Line:** During **Training**
- Overlapping Trajectories
  - **<At higher level>**
    - Inferring higher-level intention latent in observed patterns



Result
Neural Activation while Minimizing Visual Prediction Error

- **Trajectory of Neural Activation**
  - **Dashed Lines:** During **PE Minimization**
  - **Solid Line:** During **Training**
- Overlapping Trajectories
  - **<At higher level>**
    - Inferring higher-level intention latent in observed patterns
  - **<At lower level>**
    - Recalling the corresponding representations
    - Retrieval of missing sensorimotor signals



"Predictive Coding Account of MNS" (Kilner et al., 2007)

Conclusion

- Predictive Visuo-Motor Deep Dynamic Neural Network (P-VMDNN)
  - Builds a **Predictive Internal Model** of the environment
    - From dynamic sensorimotor experience
  - **Mentally Simulates** possible outcome of an action
    - With a given intention through the top-down mechanism
  - **Minimizes PE** through updating internal states
    - Inferring higher-level intention latent in observed patterns
    - Recalling the corresponding visuo-proprioceptive representations acquired during training


Jungsik Hwang


Jinhyung Kim


Ahmadreza Ahmadi


Minkyu Choi


Jun Tani

KAIST

OIST

THANK YOU

- Acknowledgement
  - Korea's ICT R&D program of MSIP/IITP. [2016(R7117-16-0163), Intelligent Processor Architectures and Application Software for CNN-RNN]
  - SRS (Special Research Student) program at OIST