Simulating Fast-Movements with Mixture-of-Expert models

An Interactive Boxing Controller

Chirag Bhuvaneshwara

Supervisor:
Janis Sprenger

Reviewers:
Prof.Dr.-Ing. Philipp Slusallek
Dr. Fabrizio Nunnari

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

UNIVERSITÄT DES SAARLANDES

Agenda

- Introduction
- Related Work
- Data Acquisition
- Boxing Neural Model
- Experiments
- Discussion

Motivation

Mortal Kombat X:

Limited punch actions

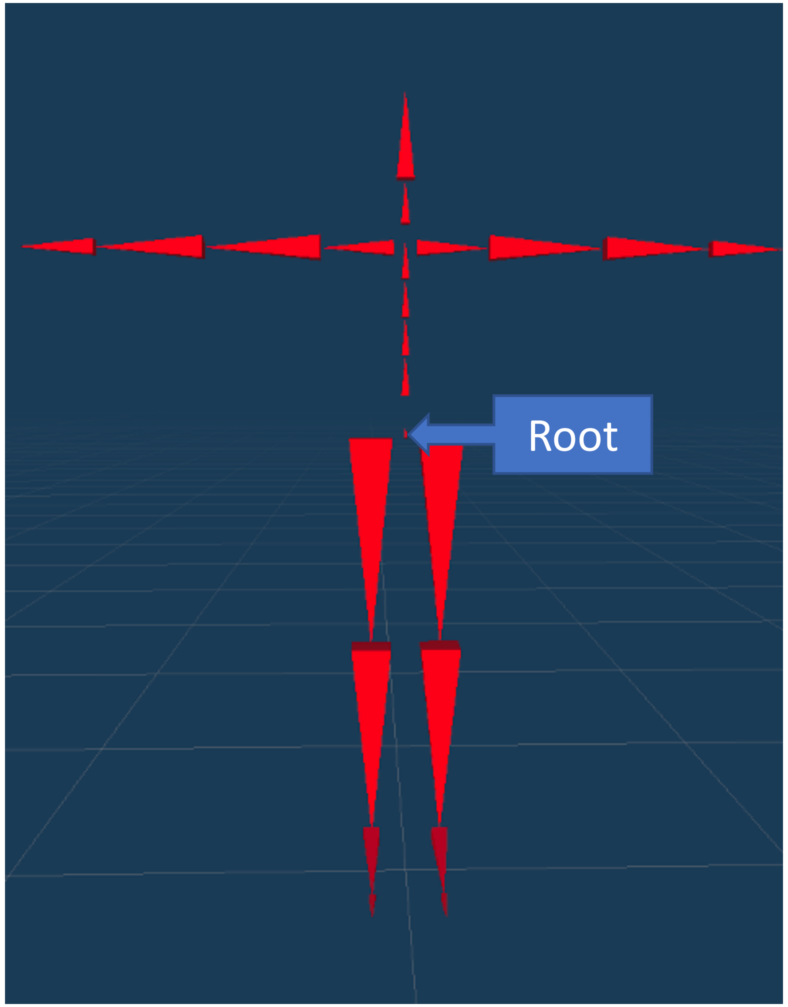
GTA 5:

Punch missing

Synthesise targeted punches
from boxing!

What is motion synthesis?

- ❑ Virtual World
- ❑ Virtual skeleton
  - Hierarchy
  - Parent of all ➔ Root joint
- ❑ Manipulate skeleton



Use-cases

- ❑ Computer games
- ❑ Animated movies

Approaches

- ❑ Kinematics
  - Use position, velocity, acceleration
- ❑ Dynamics
  - Apply rules of Physics

Challenges in boxing synthesis

- ❑ Boxing actions are fast
- ❑ Boxing actions are non-periodic
- ❑ Targeted punch control
  - Realism
  - Responsiveness
- ❑ Stepping has distinct style

Boxing specific

- ❑ Dynamic models bring in more complexity
- ❑ Real-time user interaction

Combo is too complicated!

Use kinematic model first

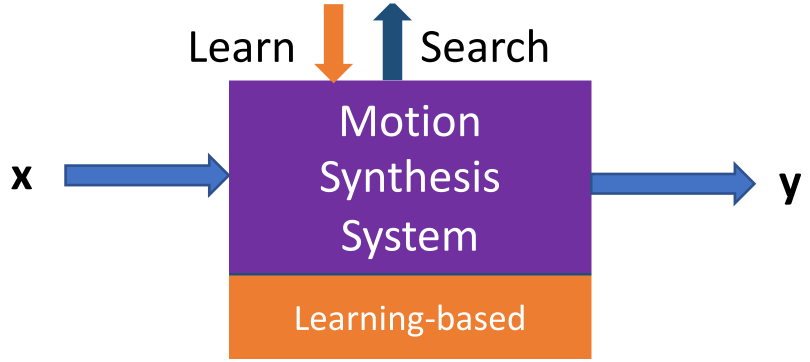
Data-driven motion synthesis

- Reference motion data

- Categorized into:
  - Search-based
  - Learning-based

- Search-based
  - Database required
  - Scalability issue

- Learning-based
  - Function replaces database
  - No search required

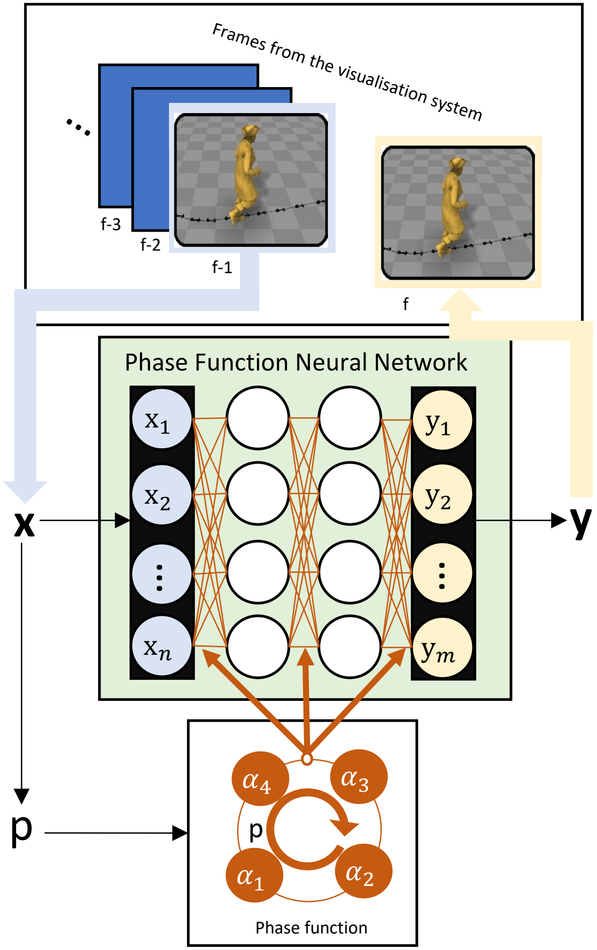


Use learning-based approach

Phase Function Neural Network (PFNN)

(Holden et al. 2017)

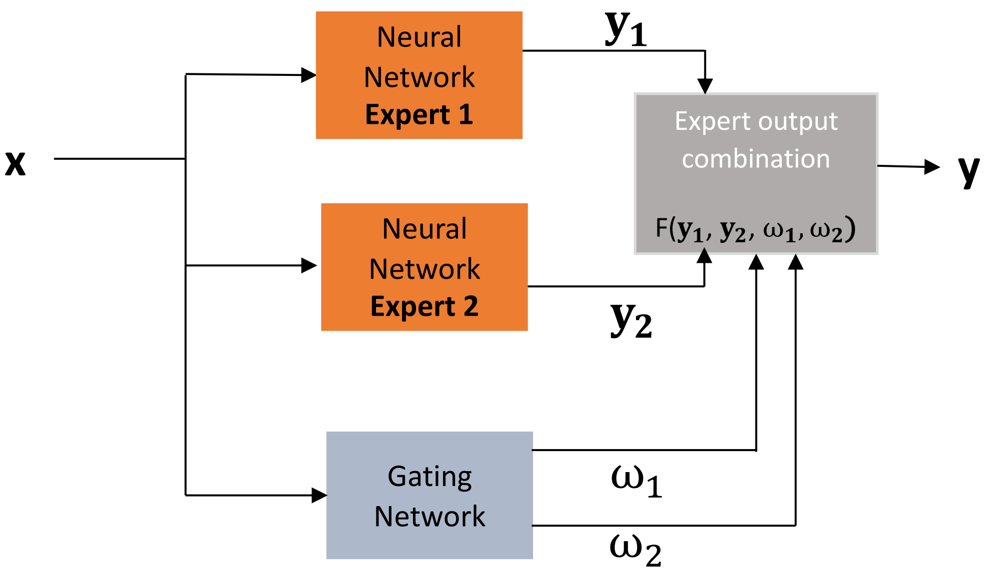
- ❑ Locomotion tasks
- ❑ Change network weights dynamically
- ❑ Fixed phase function



Vis images from Holden et al. (2017)

Mixture of Experts
(MoE)

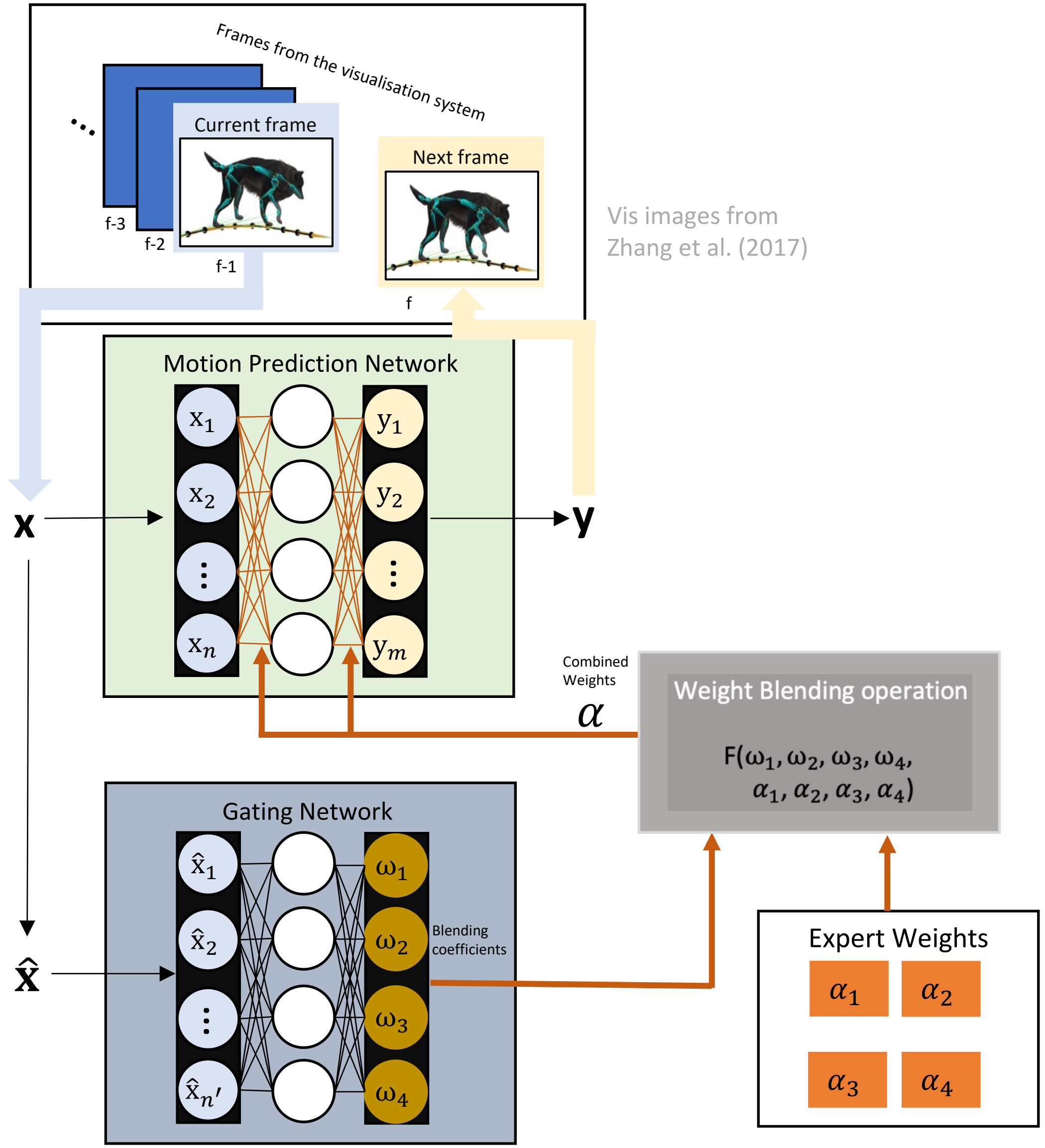
- Zhang et al. (2017)
    - Qudruped control
    - Gait labelling hard
    - PFNN not sufficient
- Mixture of Experts
    - Developed by Jacobs et al. (1991)
    - Neural network experts
    - Gating network
    - Each expert learns one action



Mode Adaptive Neural Network (MANN)

(Zhang et al. 2017)

- ❑ PFNN ➔ Change network weights dynamically
- ❑ Mixture of Experts ➔ Have multiple experts
- ❑ MANN
  - Motion Prediction Network
  - Gating Network
  - Set of Expert weights!



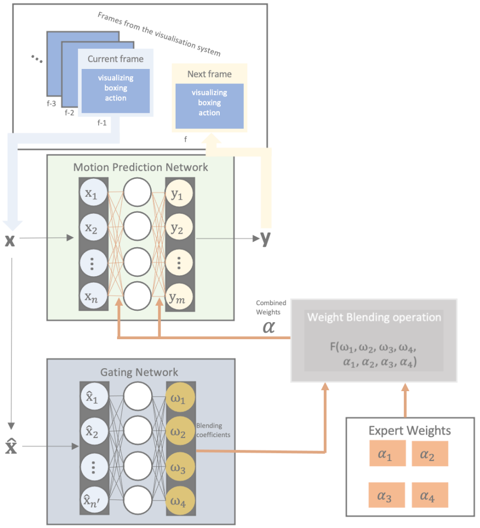
Vis images from Zhang et al. (2017)

Subsequent MoE models

- ❑ Starke et al. 2019
  - Humanoid locomotion
  - Interaction with objects
- ❑ Starke et al. 2020
  - One-on-one basketball
  - Interaction with opponent
- ❑ Starke et al. 2021
  - Mixed martial arts
  - Punching and kicking

MoE models for boxing?

- ❑ Boxing actions are non-periodic
  - ▪ Like jumping (in Zhang et al. (2017))
- ❑ Boxing stepping
  - ▪ Similar to humanoid locomotion (Starke et al. (2019, 2020, 2021))
- ❑ Additional challenges
  - ▪ Targeted punch control
  - ▪ Stepping in boxing style

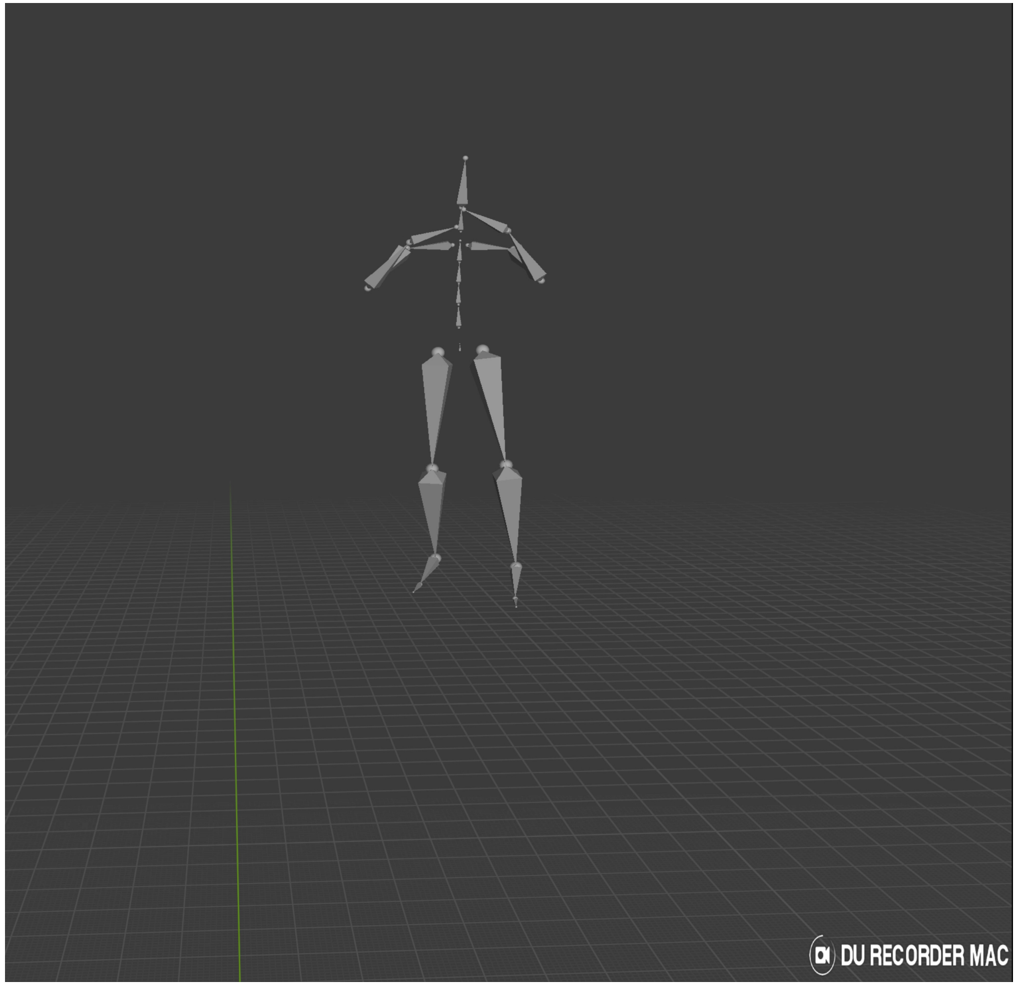


Setup

- ❑ Xsens Awinda Mocap system
  - 60 Hz ➔ 16.7 ms
- ❑ Single actor
  - Intermediate kickboxing exp
- ❑ Capturing region
  - Similar to boxing ring
- ❑ Capturing instructions
  - Shadow boxing
    - Stepping
    - Punching

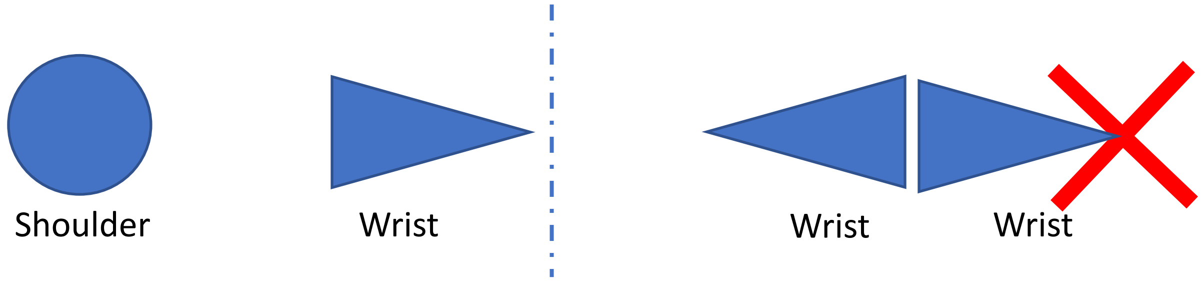
Cleanup & Result

- ❑ Cleanup
  - Removing unnecessary data
  - Fixing error from sensor slippage
  - Foot skating correction
- ❑ 10 mins of HQ data

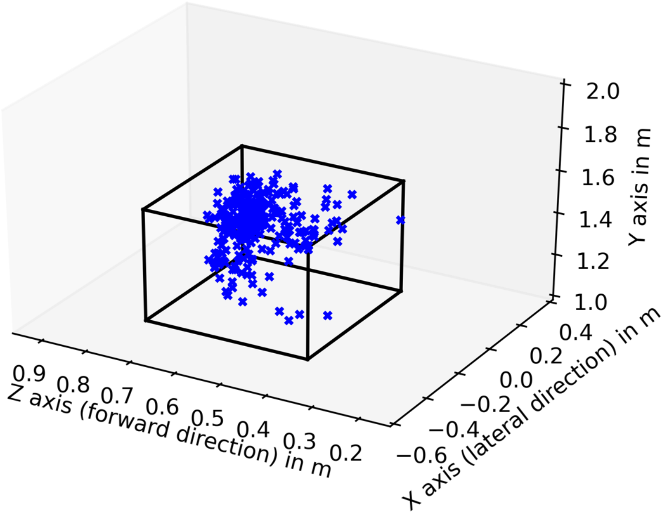


Annotation & Statistics

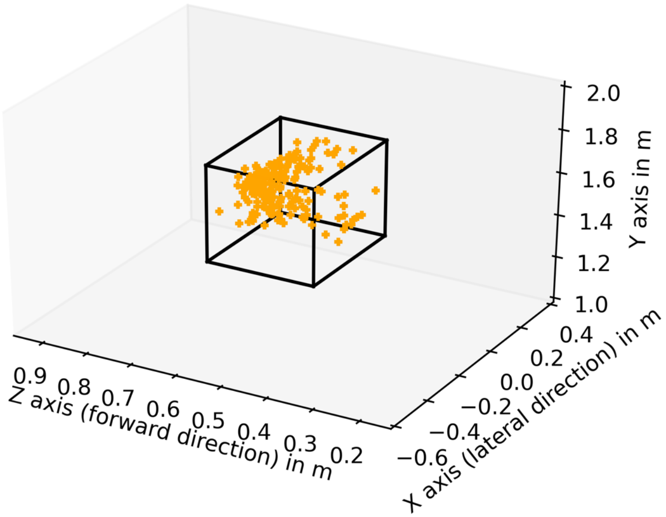
- ❑ Automatic punch annotation
- ❑ Targets in local space of skeleton
  - Ground projected root at O
  - Looking at +Z axis
  - Root transformation



Dataset left punch targets



Dataset right punch targets



Punch type	Num punches	Mean duration (frames)
Overall	451	27

Step type	Mean duration (frames)
Overall	20

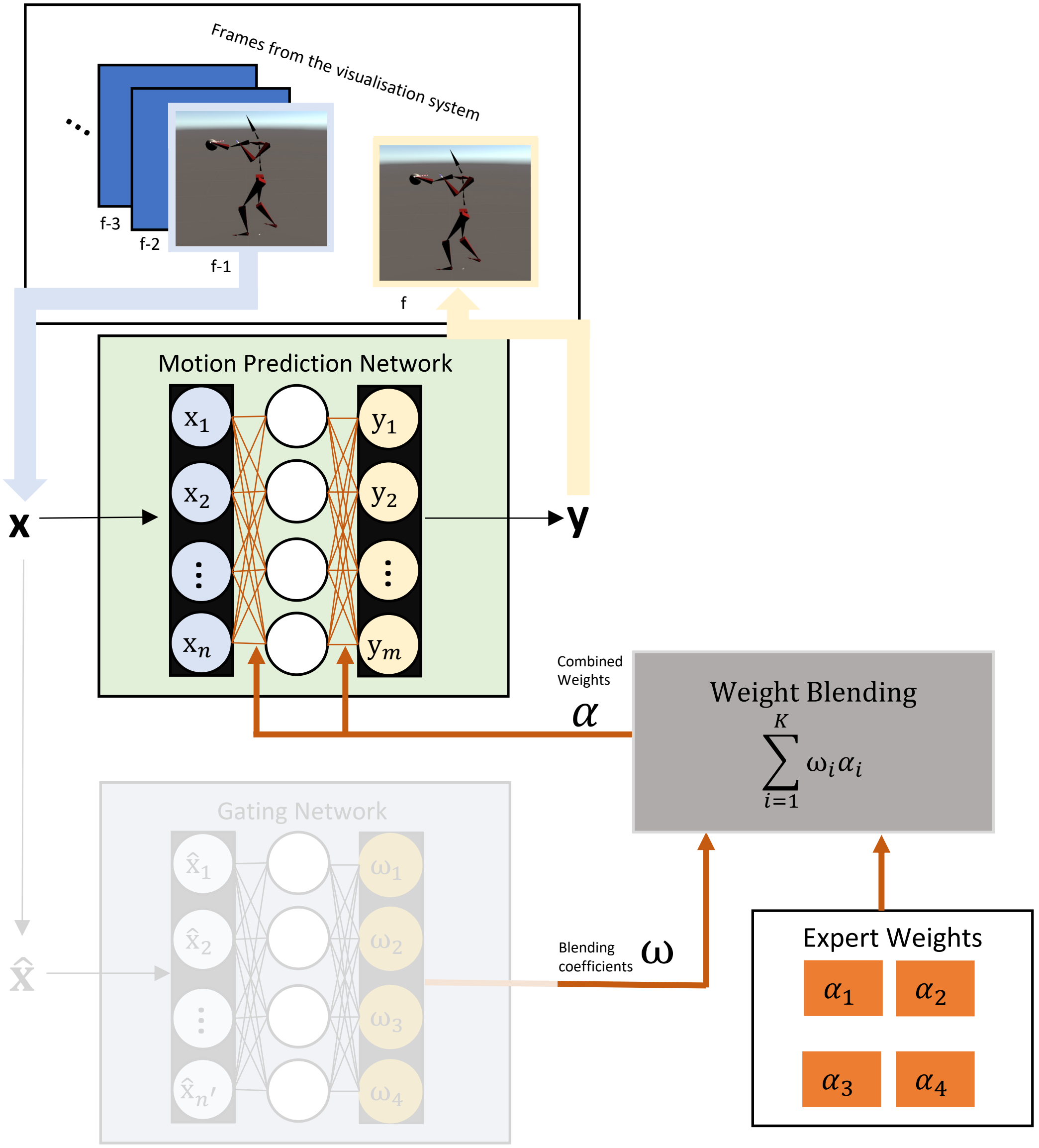
Motion Prediction Network (MPN)

- 3 layers
- Equation

$$\boldsymbol{\Theta}(\mathbf{x}, \boldsymbol{\alpha}) = \boldsymbol{W_2}\mathrm{ELU}(\boldsymbol{W_1}\mathrm{ELU}(\boldsymbol{W_0 x} + \boldsymbol{b_0}) + \boldsymbol{b_1}) + \boldsymbol{b_2}$$

- Weight blending

$$\sum_{i=1}^{K} \omega_i \alpha_i$$



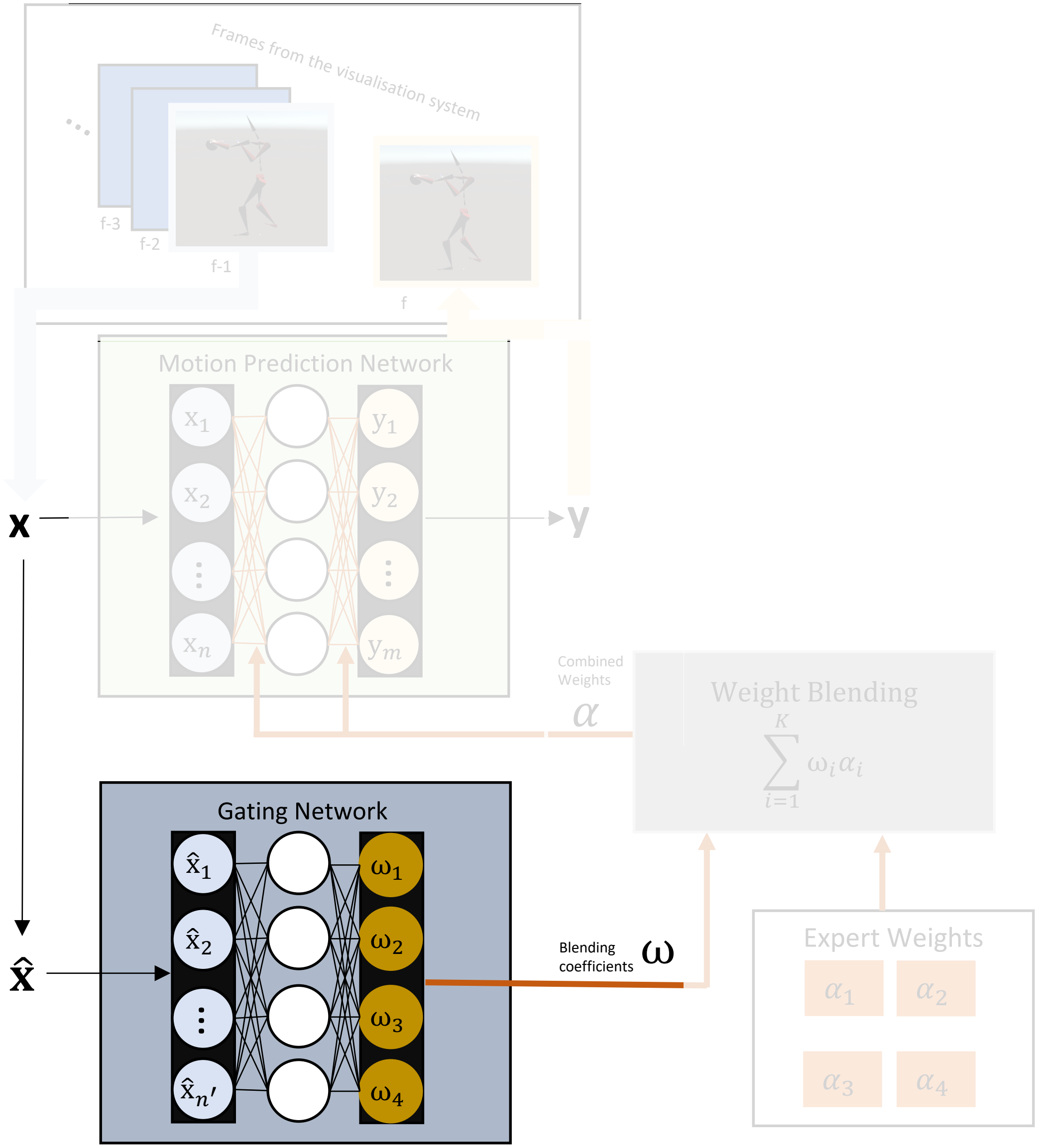
Gating Network
(GN)

- 3 layers
- Equation

$$\boldsymbol{\Omega}(\hat{\mathrm{x}}, \boldsymbol{\mu}) = \sigma(\hat{\boldsymbol{W}}_{\mathbf{2}}\mathrm{ELU}(\hat{\boldsymbol{W}}_{\mathbf{1}}\mathrm{ELU}(\hat{\boldsymbol{W}}_{\mathbf{0}}\hat{\boldsymbol{x}} + \boldsymbol{b}'_{\mathbf{0}}) + \boldsymbol{b}'_{\mathbf{1}}) + \boldsymbol{b}'_{\mathbf{2}})$$

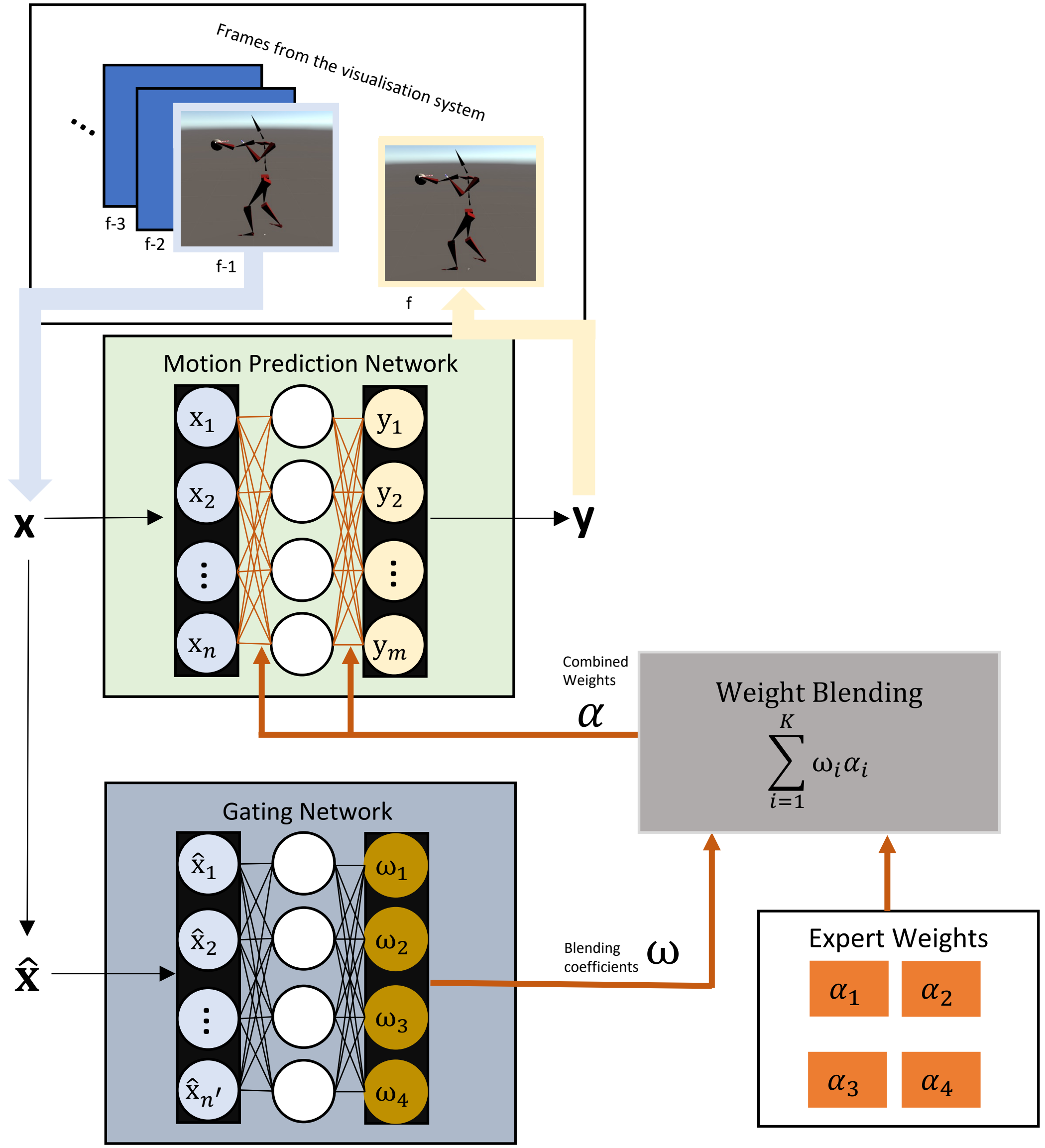
- Activation function

$$\mathrm{ELU(z)} = \max(\mathrm{z}, 0) + \exp(\min(\mathrm{z}, 0)) - 1$$



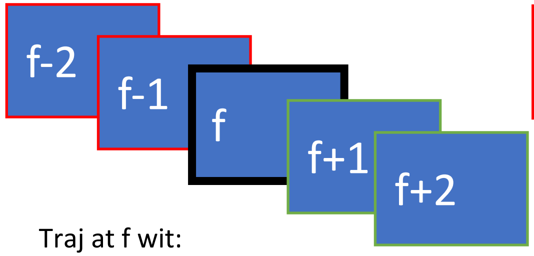
Training Details

- Training data: $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots]$ $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_\mathbf{1}, \hat{\mathbf{x}}_\mathbf{2}, \ldots]$
  $\mathbf{Y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots]$
- MSE loss function:
  $$Cost(\mathbf{X}, \mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\mu}) = \|\mathbf{Y} - \boldsymbol{\Theta}(\mathbf{X}, \boldsymbol{\Omega}(\hat{\mathbf{X}}; \boldsymbol{\mu}); \boldsymbol{\beta})\|_2^2$$
- Batch size = 32
- Optimization: Adam with warm restarts
- Training time:
  30 – 120 mins on NVIDIA Quadro RTX 8000



Motion States

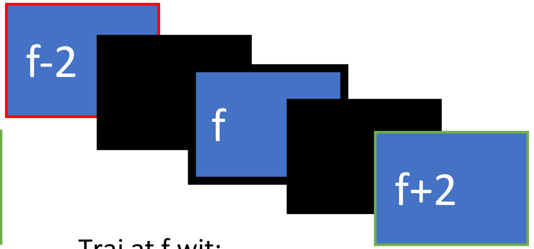
- Local space of skeleton
  - Rotate to face +Z axis
  - Place O at ground projected root
  - Root transformation
- Trajectory



Traj at f wit:
- window size w = 5
- Frame skip fs = 1

$$Tr_{pos} = [\; pos_{f-2},\; pos_{f-1},\; pos_f,\; pos_{f+1},\; pos_{f+2} \;]$$

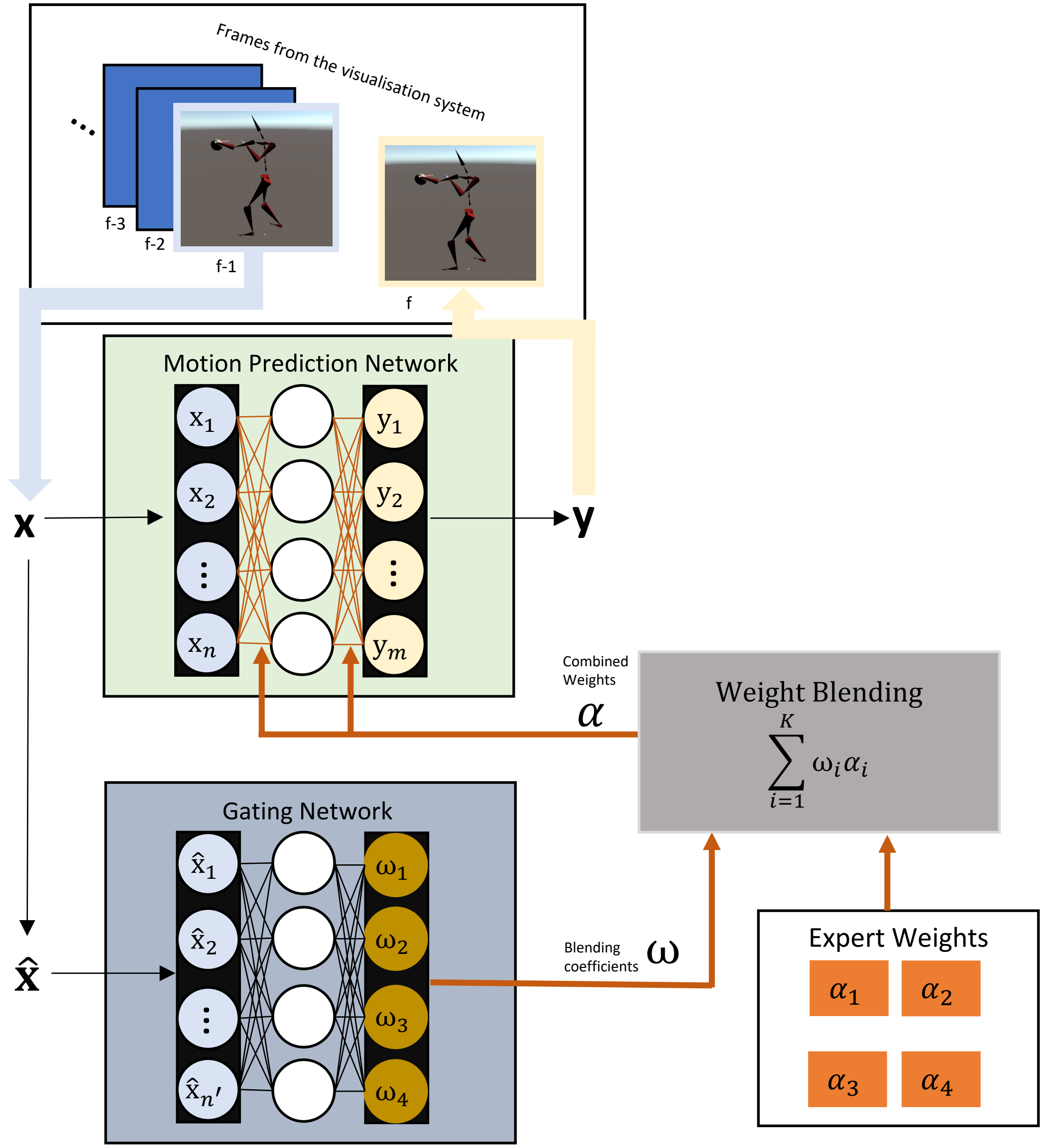
Past (pos_{f-2}, pos_{f-1})
Current (pos_f)
Future (pos_{f+1}, pos_{f+2})



Traj at f wit:
- window size w = 3
- Frame skip fs = 2

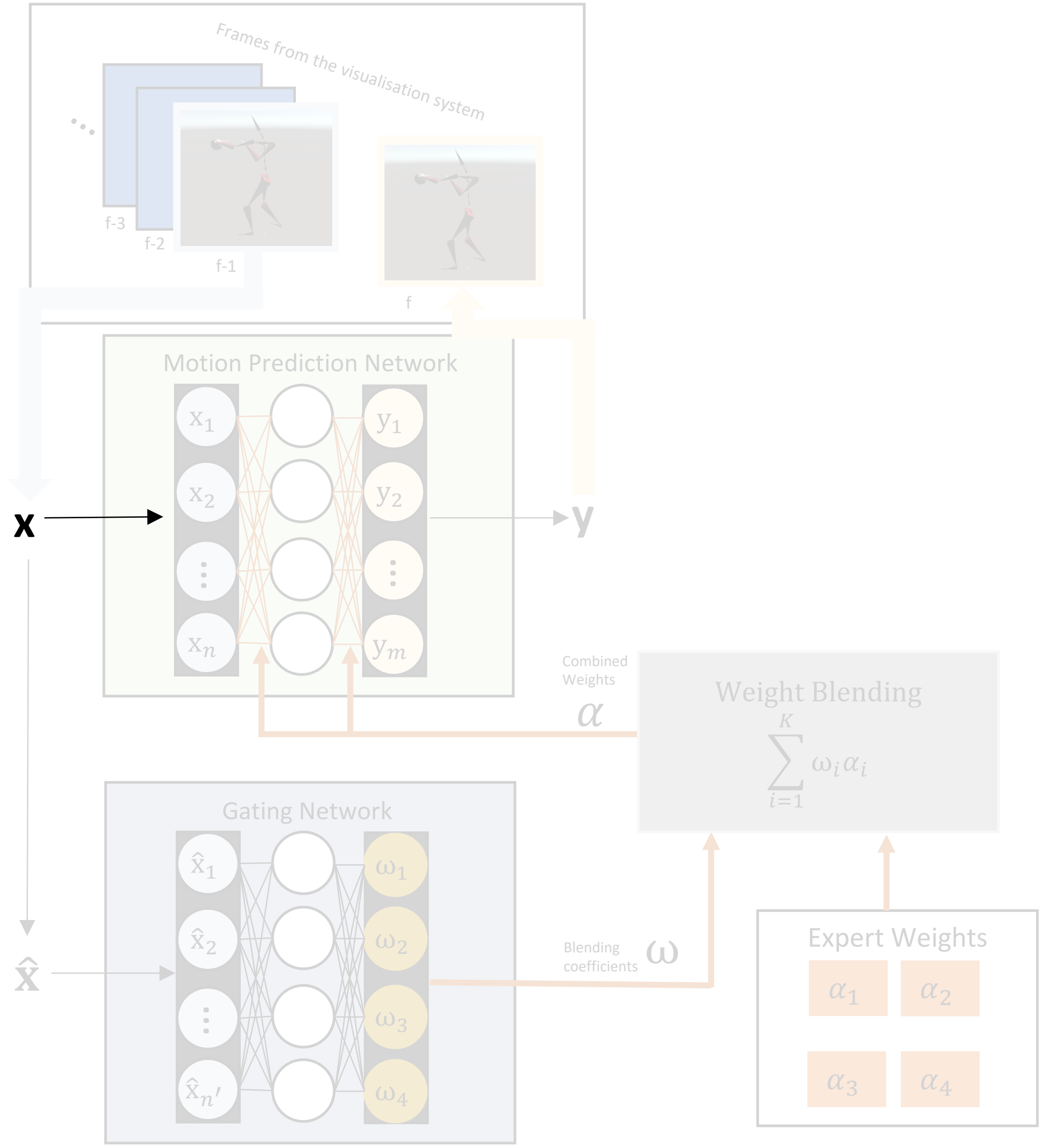
$$Tr_{pos} = [\; pos_{f-2},\; pos_f,\; pos_{f+2} \;]$$

Past (pos_{f-2})
Current (pos_f)
Future (pos_{f+2})



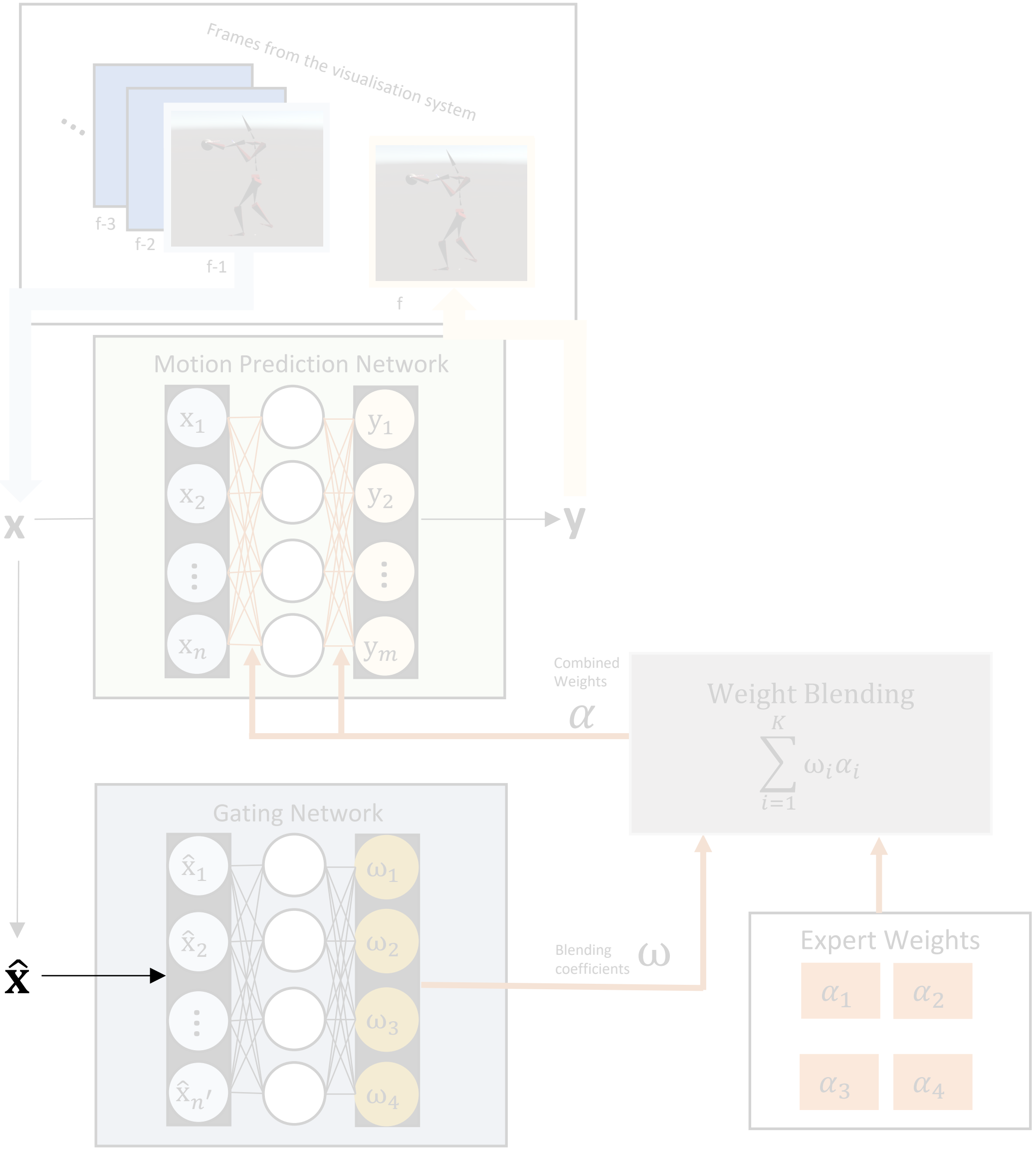
MPN Input

$$
\begin{aligned}
\mathbf{x_f} = \{ & \left.\begin{array}{l}\mathbf{t_f^{rp}},\\ \mathbf{t_f^{rv}},\\ \mathbf{t_f^{rd}},\end{array}\right\}\text{Root trajectory}\\
& \left.\begin{array}{l}\mathbf{t_f^{wp}},\\ \mathbf{t_f^{wv}},\end{array}\right\}\text{Wrist trajectory}\\
& \left.\begin{array}{l}\mathbf{a_{f-1}^{l}},\\ \mathbf{a_{f-1}^{pt}},\end{array}\right\}\text{Punch action indicators}\\
& \left.\begin{array}{l}\mathbf{j_{f-1}^{p}},\\ \mathbf{j_{f-1}^{v}}\end{array}\right\}\text{All joints}\\
& \} \in \mathbb{R}^n
\end{aligned}
$$



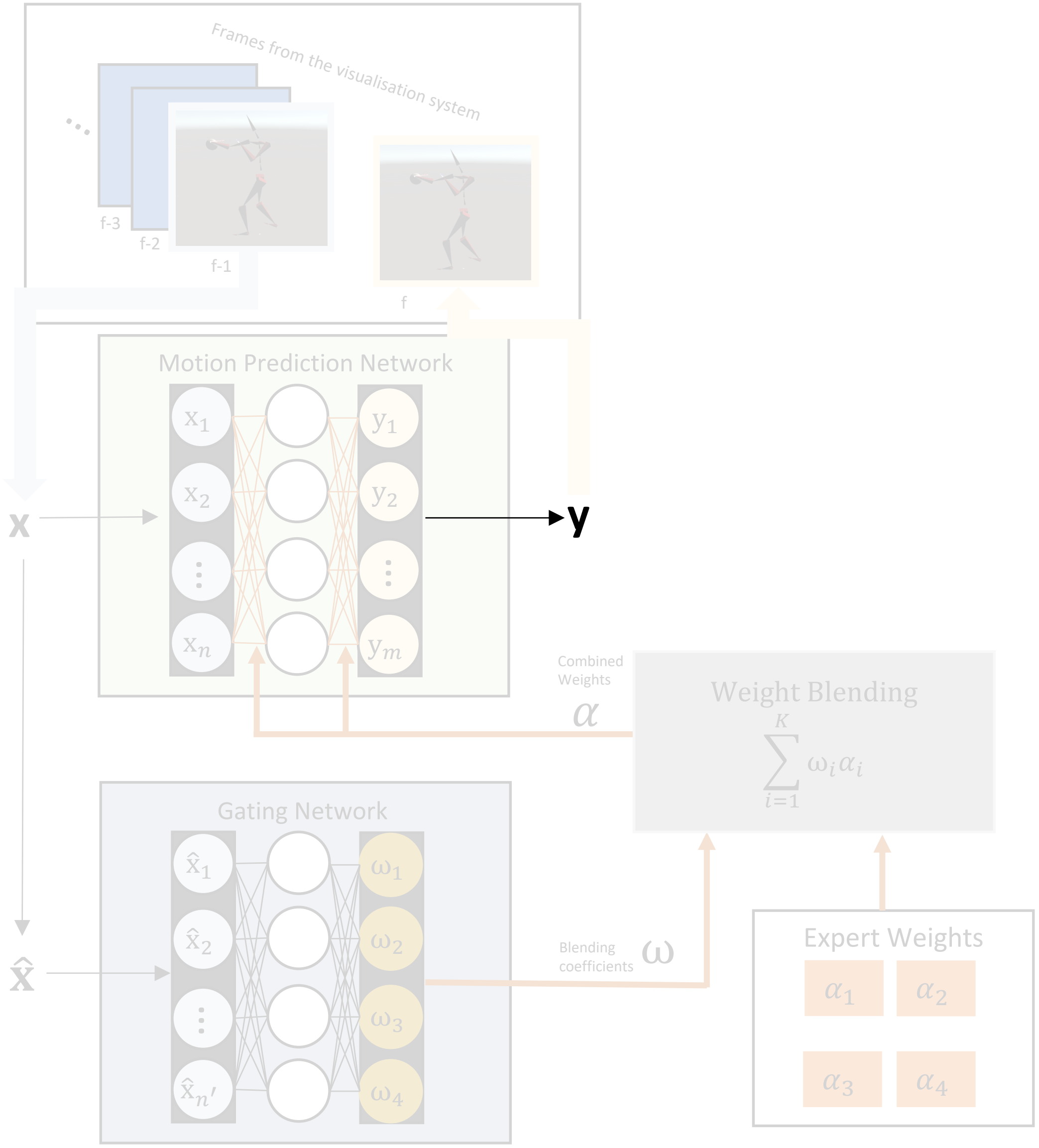
GN Input

$$\hat{\mathbf{x}}_{\mathbf{f}} = \{ \underbrace{\mathbf{t}_{\mathbf{f}}^{\mathbf{rp}}, \mathbf{t}_{\mathbf{f}}^{\mathbf{rd}},}_{\text{Root trajectory}} \underbrace{\mathbf{t}_{\mathbf{f}}^{\mathbf{wv}},}_{\text{Wrist trajectory}} \underbrace{\mathbf{h}_{\mathbf{f-1}}^{\mathbf{efv}},}_{\text{Wrist velocity}} \underbrace{\mathbf{a}_{\mathbf{f-1}}^{\mathbf{l}},}_{\text{Punch action indicator}} \underbrace{\mathbf{b}_{\mathbf{f-1}}^{\mathbf{efv}}, \mathbf{b}_{\mathbf{f-1}}^{\mathbf{efp}}}_{\text{Foot joint}} \} \in \mathbb{R}^{n'}$$



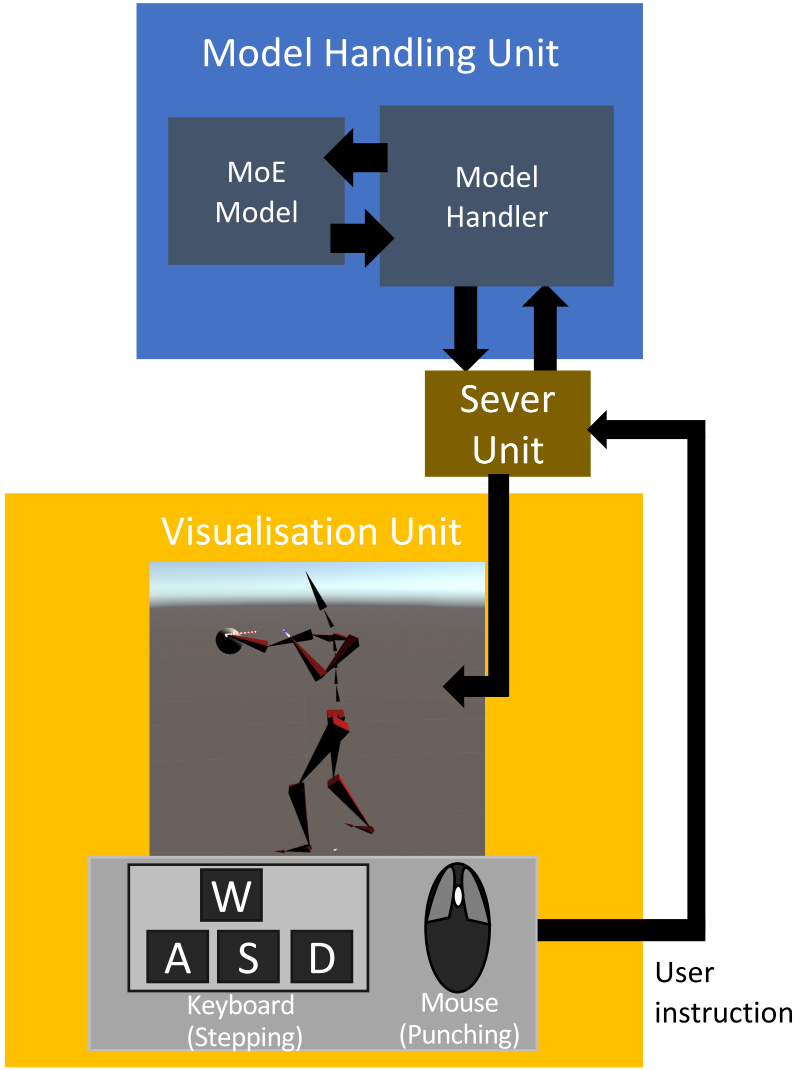
MPN Output

$$\mathbf{y_f} = \{ \underbrace{\mathbf{t^{rp}_{f+1}}, \mathbf{t^{rv}_{f+1}}, \mathbf{t^{rd}_{f+1}}}_{\text{Root trajectory}}, \underbrace{\mathbf{t^{wp}_{f+1}}, \mathbf{t^{wv}_{f+1}}}_{\text{Wrist trajectory}}, \underbrace{\mathbf{a^{l}_{f}}, \mathbf{a^{pt}_{f}}}_{\text{Punch action indicators}}, \underbrace{\mathbf{q^{p}_{f}}, \mathbf{q^{d}_{f}}}_{\text{Root joint}}, \underbrace{\mathbf{b^{c}_{f-1}}}_{\text{Foot contact}}, \underbrace{\mathbf{j^{p}_{f}}, \mathbf{j^{v}_{f}}}_{\text{All joints}} \} \in \mathbb{R}^m$$



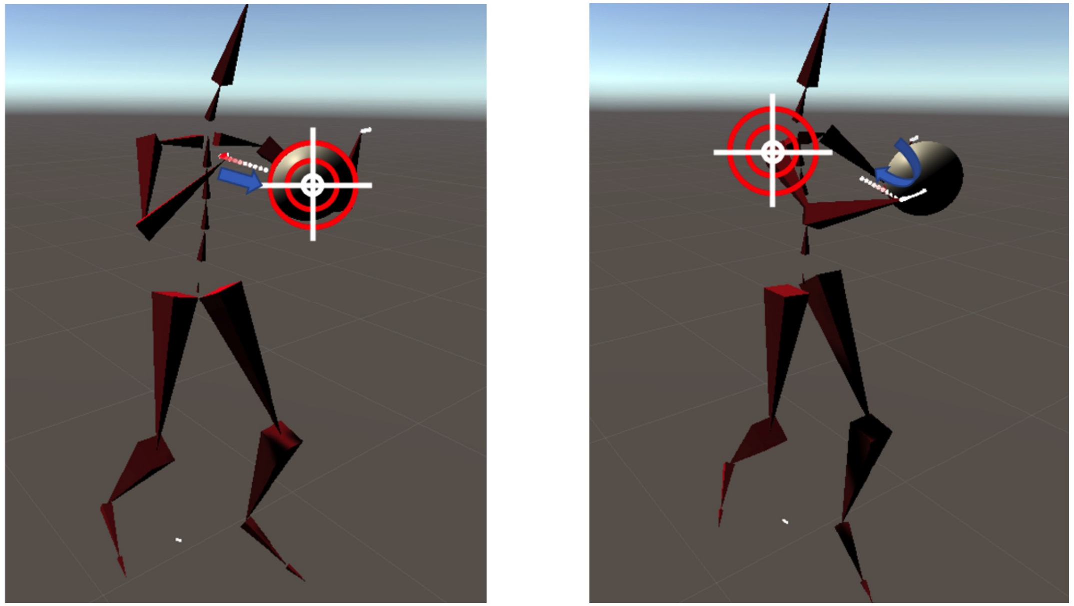
Boxing Controller System

- ❑ Visualisation unit
  - User input
  - Motion visualisation
- ❑ Model Handling Unit
  - Root transformation
  - Merge user input
- ❑ Server Unit

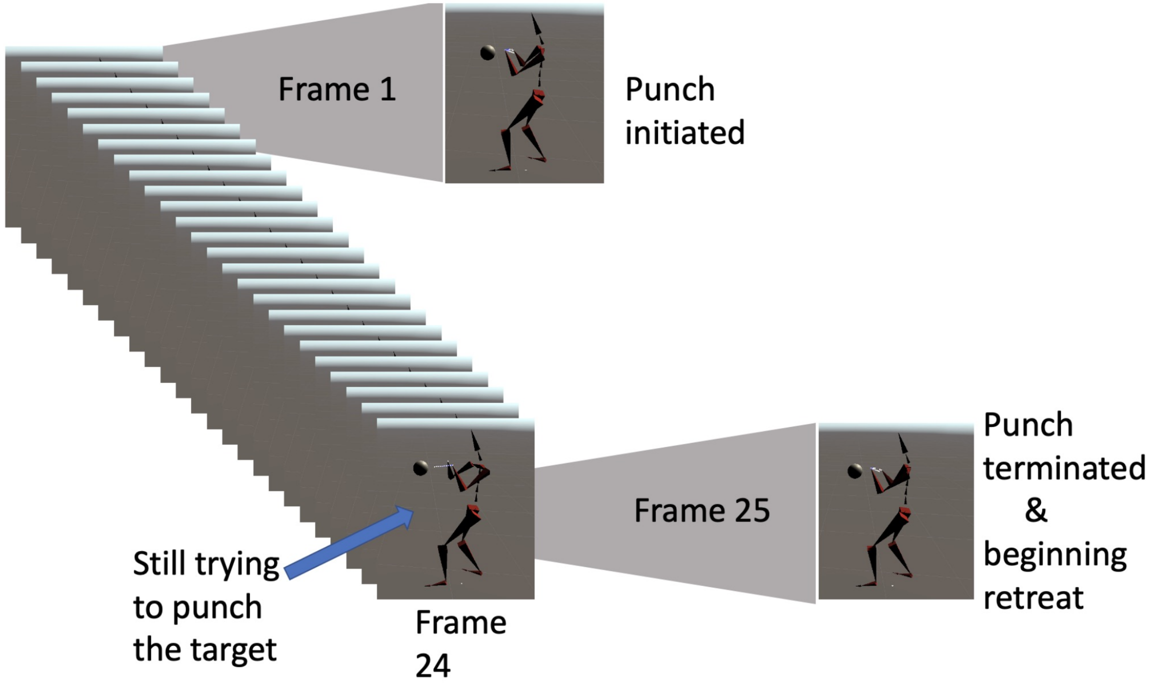


Boxing Controller System

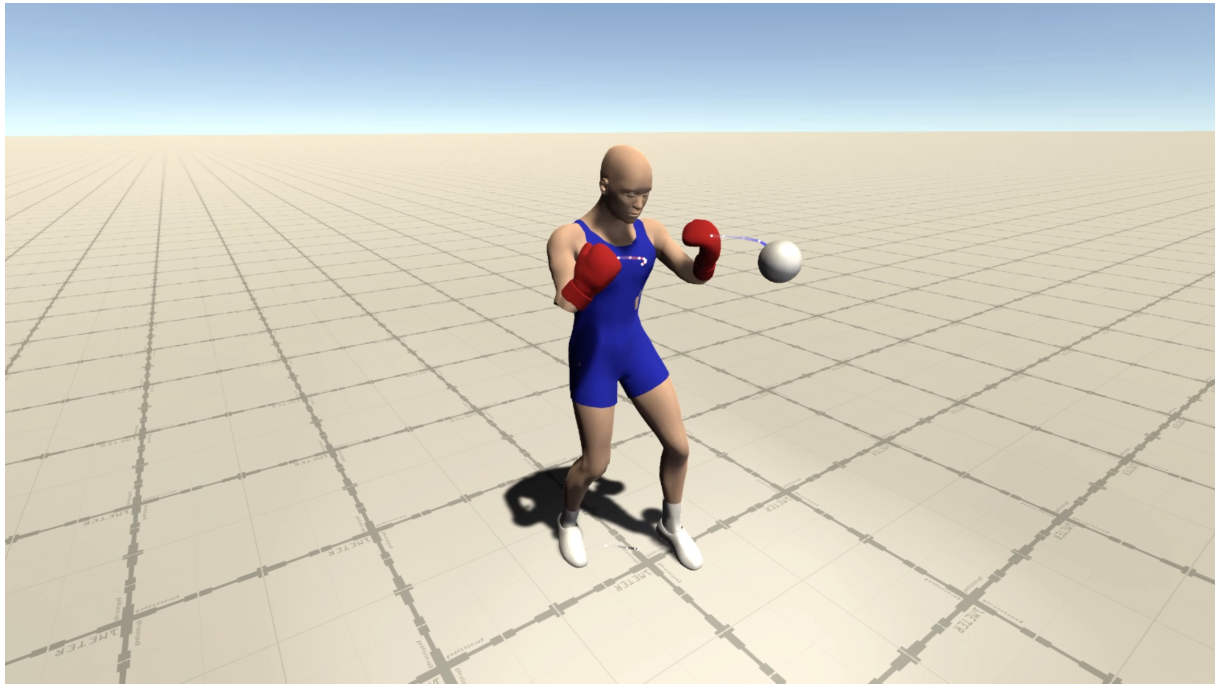
Punch target management

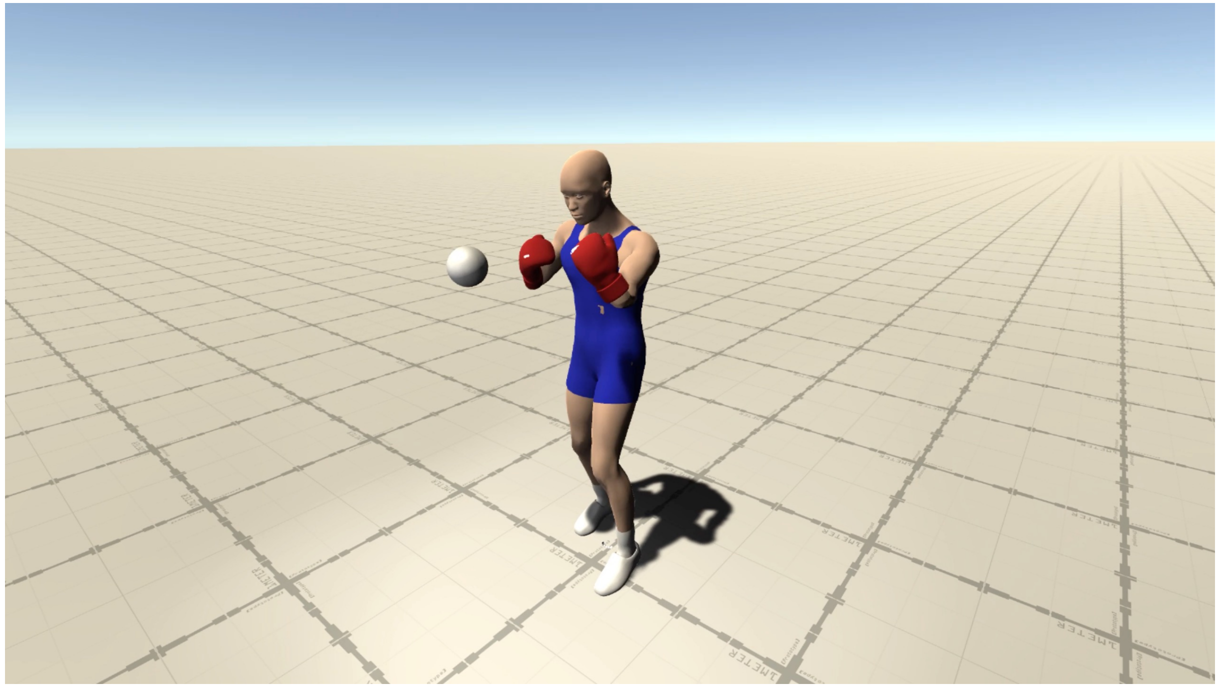


Punch termination on miss

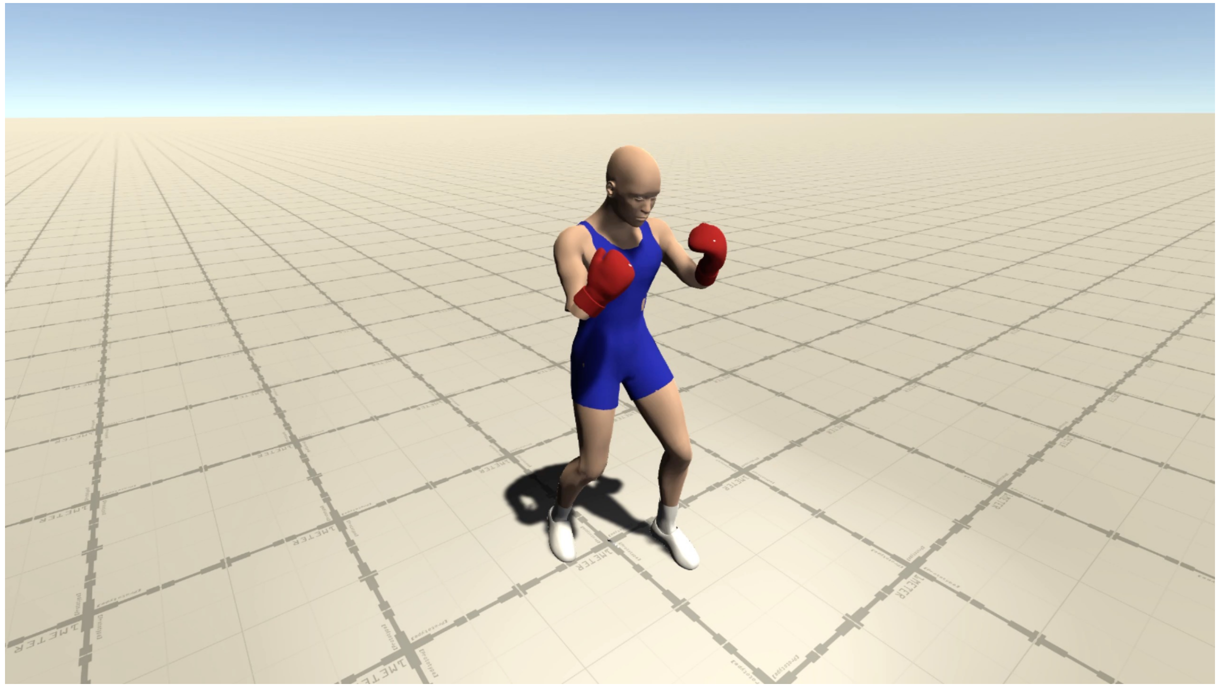


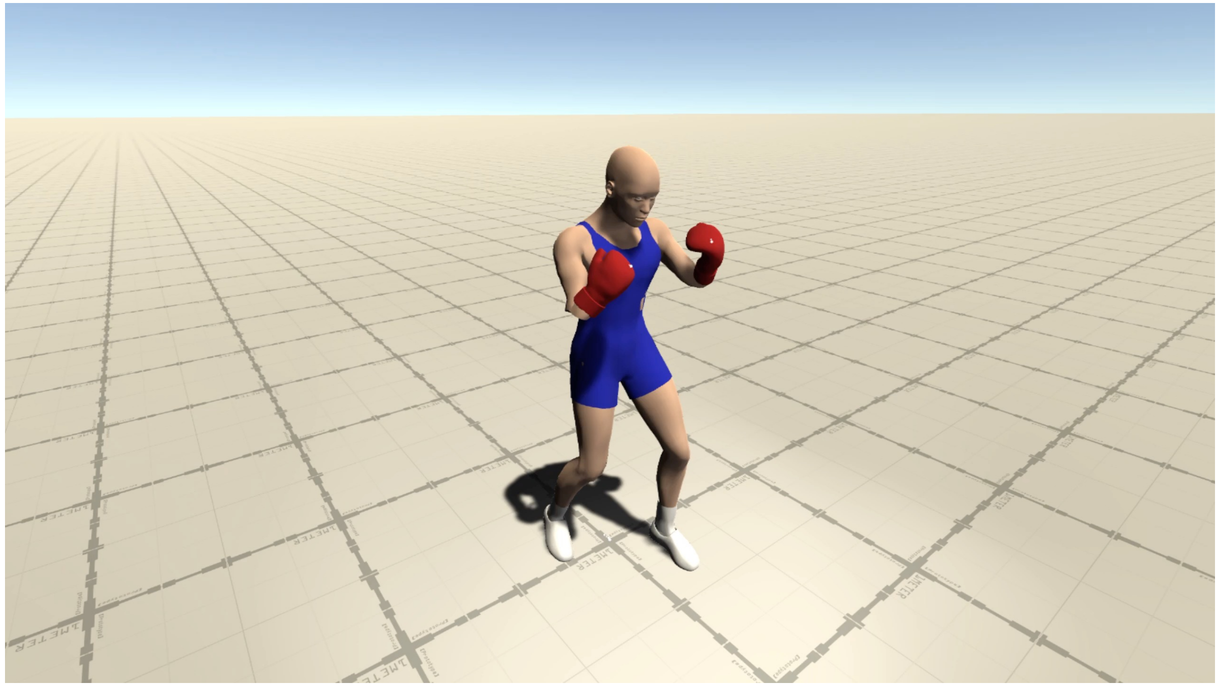
Visualisation: Punching



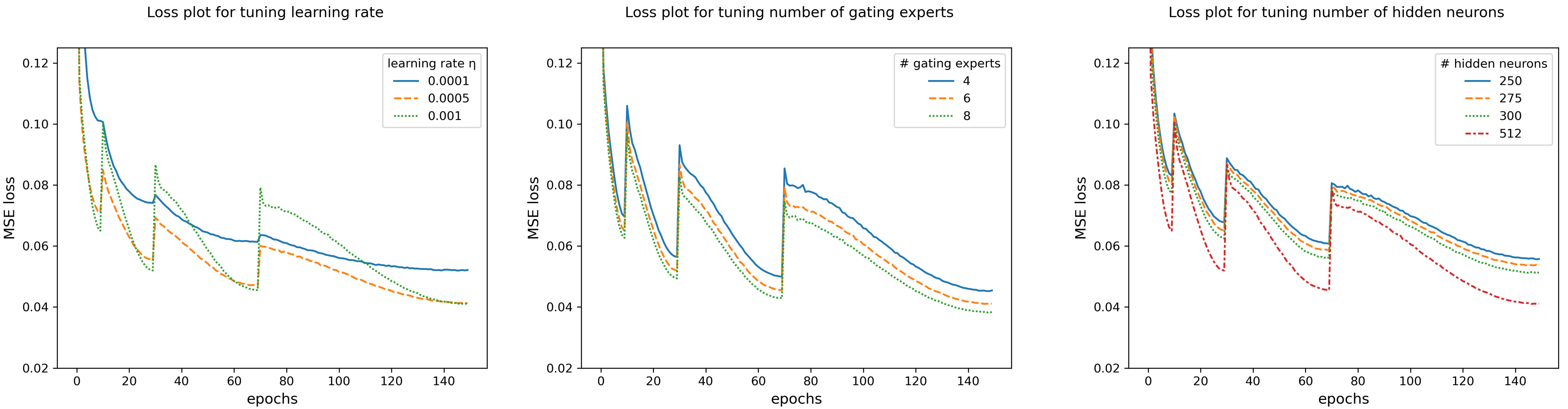


Visualisation: Stepping





Hyperparameter Tuning



❑ Best parameters:

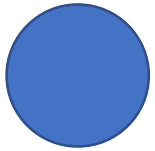
- Learning rate = 0.001
- Num expert weights = 8
- Num MPN hidden neurons = 512

Punch control
Evaluation Metrics

- Punch Accuracy: How many punches are good enough?

$$accuracy = \begin{cases} 1; & \mathbf{l^2}(x,y) < p_{threshold} \\ 0; & \mathbf{l^2}(x,y) \geq p_{threshold} \end{cases}$$

$$\mathrm{PA} = \frac{B}{Z} \times 100$$



Avergae German male head

- Punch Error: How close do we get to target?

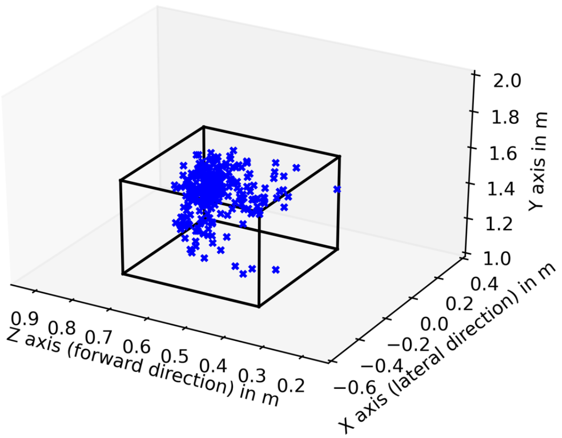
$$\mathrm{PE}(\mathbf{W}, \mathbf{P}) = \frac{1}{B} \sum_{n=1}^{B} \mathbf{l^2}(\mathbf{w}_n, \mathbf{p}_n)$$

- Foot Skating Error: Does the character float?

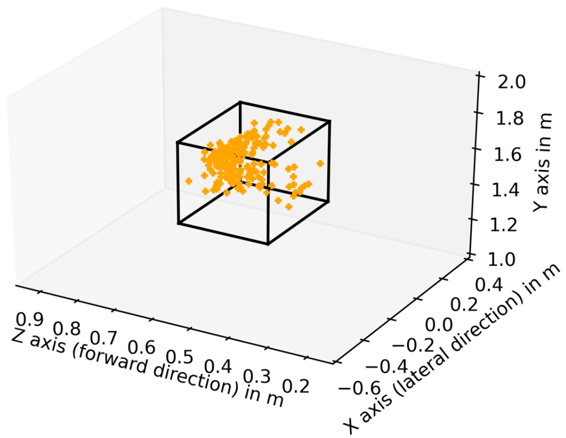
$$\mathrm{FSE} = \frac{1}{2} \sum_{l=1}^{2} d_l \left(2 - 2^{\left(\frac{h_l}{H}\right)}\right)$$

Punch control
Setup

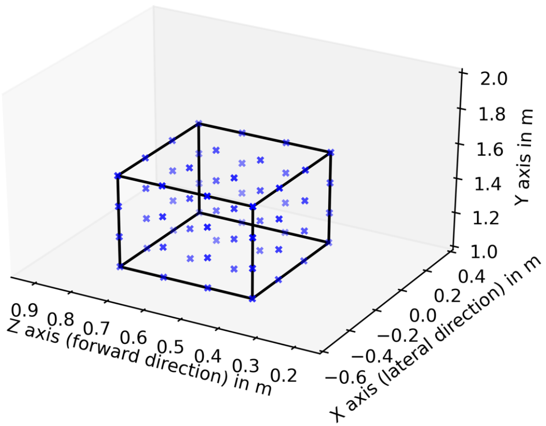
Dataset left punch targets



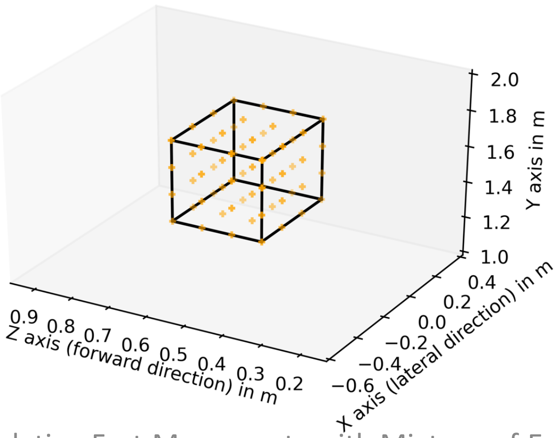
Dataset right punch targets

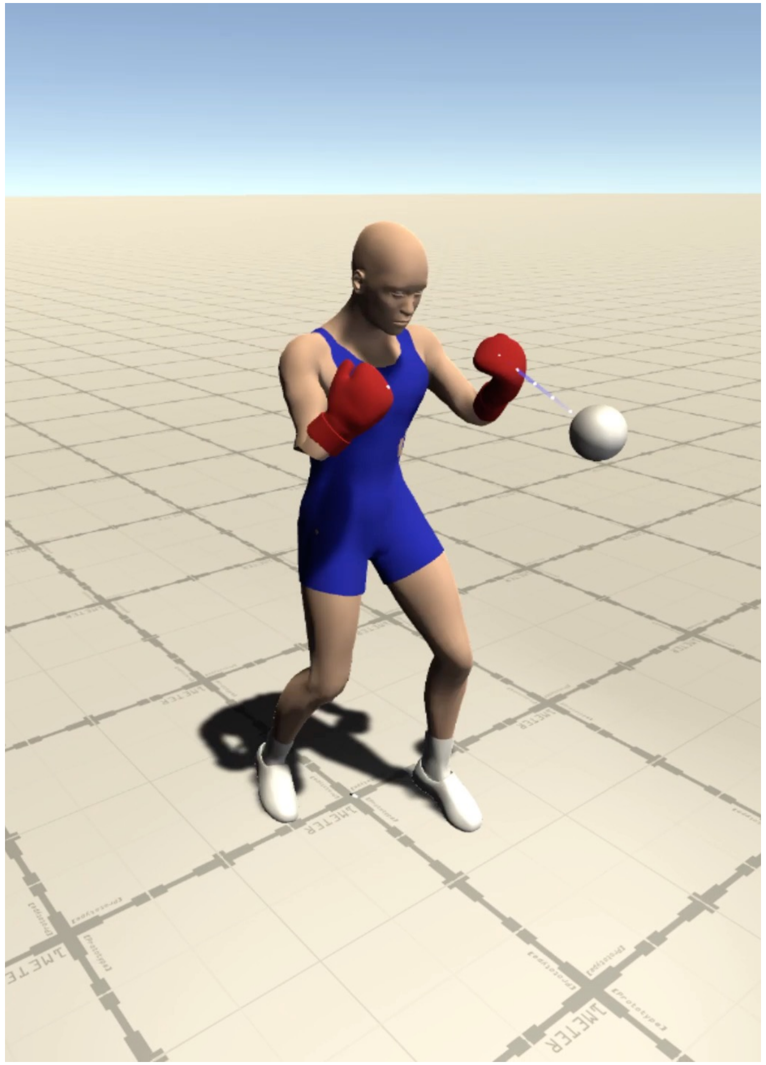


Uniform left punch targets



Uniform right punch targets





Punch control
Trajectory Sampling

- Assumption:
  - Traj similar to dataset punch
    ➔ more accurate punches
  - Models with trajectory window
    - covering approx. 27 frames (0.45s)
- Observation:
  - w=14, fs=3 leads to best accuracy.
  - Best model has somewhat high FSE.
  - Accuracy of left hand always higher than right hand.
- Conclusion:
  Sample according to dataset punches

w $(frames)$	fs $(frames)$	$PE_{overall}$ (cm)	$PA_{overall}$ (%)	FSE (cm)
4	1	11.43	20.31	**0.285**
6	1	11.29	28.91	0.305
8	1	11.02	28.91	0332
10	1	11.32	32.81	0.336
12	1	11.12	34.38	0.403
14	1	11.08	37.50	0.341
16	1	11.04	26.56	0.418
18	1	10.92	30.47	0.378
20	1	11.30	28.91	0.404
12	2	10.95	37.50	0.321
14	2	**10.20**	36.72	0.320
16	2	10.22	39.06	0.335
18	2	10.85	42.97	0.350
20	2	11.07	33.59	0.804
10	3	11.00	45.31	0.549
14	**3**	11.38	**53.91**	0.667
16	3	11.25	40.62	0.777
18	3	11.12	42.97	0.846

Punch control
Ablation Study

Wrist trajectory variables	$PE_{overall}$ (cm)	$PA_{overall}$ $(\%)$	FSE (cm)
none	**11.18**	31.25	1.177
pos	12.13	35.94	0.980
pos + vel	11.38	**53.91**	**0.667**

❑ Assumption:

- Sufficient trajectory information needed
- 3 models using w=14 and fs=3

❑ Observation:

- Fewer trajectory variables leads to more steps
- All trajectory variables correctly move the hand to punch

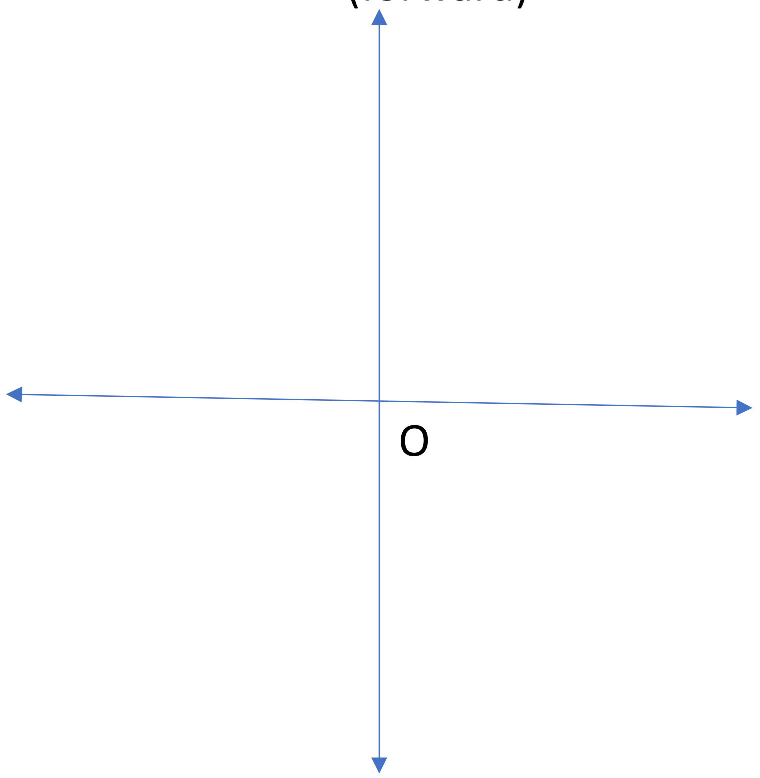
❑ Conclusion:

Traj pos + vel ➔ accurate punches

Stepping control

Setup

- ❑ Two scenarios:
  - Step forward
  - Step backward
- ❑ Stepping executed along Z axis
- ❑ Duration: 300 frames (5s)



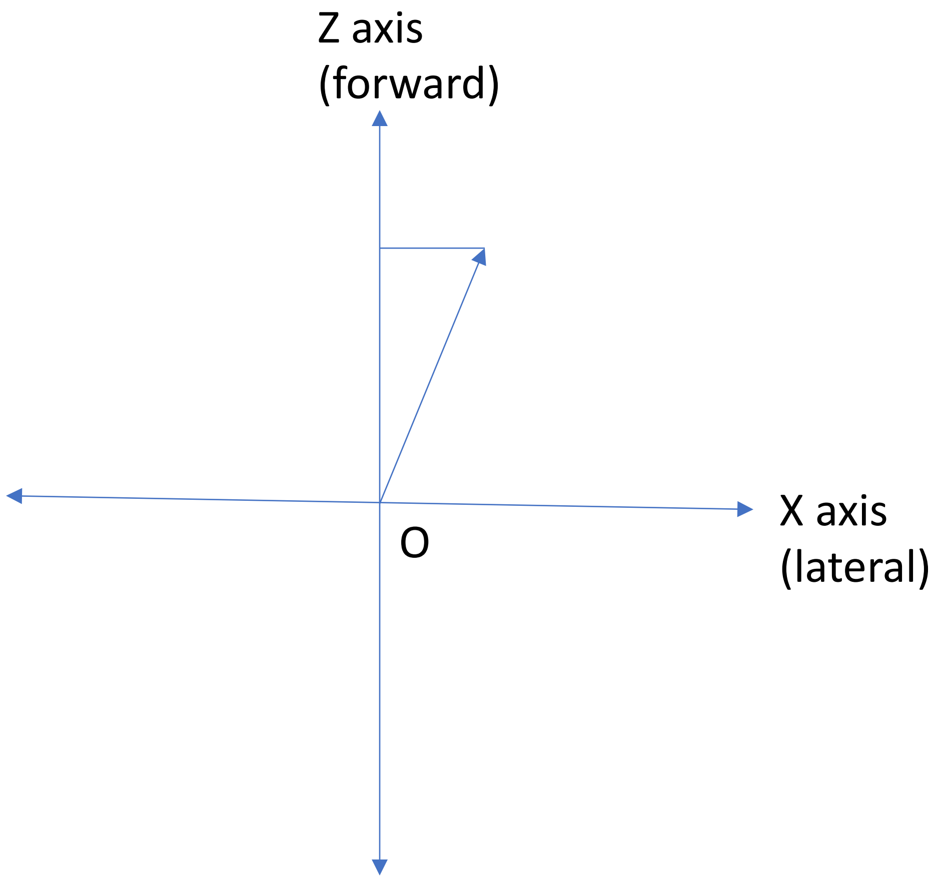
Stepping control
Evaluation Metrics

- ❑ Stepping Path Error (SPE): Path following capability

$$\mathrm{SPE}(\mathbf{G},\mathbf{E}) = \frac{1}{N}\sum_{f=1}^{N} \mathbf{l}^{\mathbf{2}}(\mathbf{g_f},\mathbf{e_f})$$

- ❑ Foot Skating Error (FSE): Does the character float?

$$\mathrm{FSE} = \frac{1}{2}\sum_{l=1}^{2} d_l \left(2 - 2^{\left(\frac{h_l}{H}\right)}\right)$$



Stepping control
Trajectory Sampling

❑ Assumption:

- Trajectory window similar to average stepping duration
  ➔ better stepping
- Trained models: 0.3s to 2s windows
- Dataset average: 20 frames (0.3s)

w $(frames)$	fs $(frames)$	$SPE_{overall}$ (cm)	FSE (cm)
10	**2**	**33.61**	**0.52**
12	2	37.36	0.70
14	2	49.75	0.76
16	2	68.92	0.56
18	2	65.23	0.97
10	3	37.84	1.28
8	4	63.42	1.36
14	4	52.14	2.59

❑ Observation:

- Forward stepping better than backward stepping

❑ Conclusion:

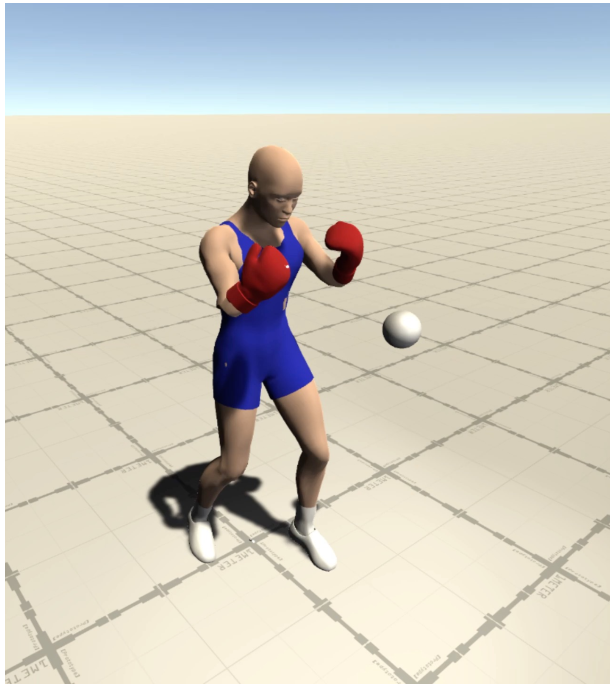
Sample according to dataset stepping action

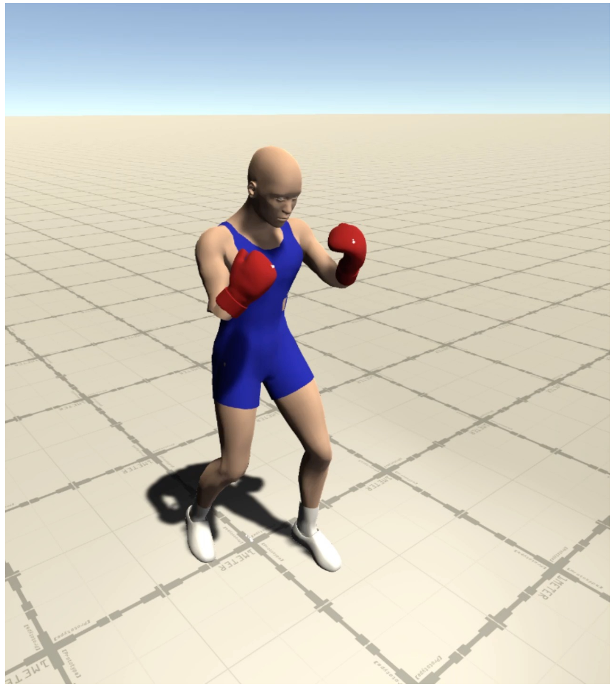
General Discussion

- ❑ Amount of data: 10 mins v/s 20 hrs
- ❑ Hyperparameter tuning led to sharper motions
- ❑ At least traj pos and vel → accurate punches
- ❑ Trajectory should be in line with dataset
- ❑ Controller fixes:
  - Punch target management
  - Punch termination

Limitations

- ❑ Less data
- ❑ Better instructions during data acquistion
- ❑ Automatic annotation
- ❑ Synthesising rotations





Future Work

- ❑ Amount of data for neural synthesis
- ❑ Update motion states
  - around boxing ring
  - around opponent
  - inlude rotations
- ❑ Other targeted actions
- ❑ Dynamic models

Conclusions

- ❑ Synthesised non-periodic, fast boxing actions
- ❑ Tune MoE models
- ❑ Trajectory variables necessary
- ❑ Trajectory sampling as per dataset
- ❑ Controller essential

References

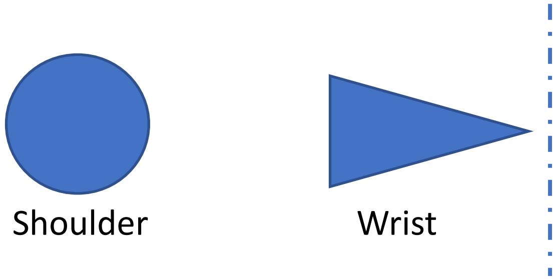
- Holden, D., Komura, T. and Saito, J., 2017. Phase-functioned neural networks for character control. *ACM Transactions on Graphics*, 36(4), pp.1-13.
- Zhang, H., Starke, S., Komura, T. and Saito, J., 2018. Mode-adaptive neural networks for quadruped motion control. *ACM Transactions on Graphics*, 37(4), pp.1-11.
- Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, Geoffrey E. Hinton; Adaptive Mixtures of Local Experts. *Neural Comput* 1991; 3 (1): 79–87.
- *Sebastian Starke, He Zhang, Taku Komura, and Jun Saito. 2019. Neural state machine for character-scene interactions. ACM Trans. Graph. 38, 6, Article 209 (November 2019), 14 pages.*
- *Sebastian Starke, Yiwei Zhao, Taku Komura, and Kazi Zaman. 2020. Local motion phases for learning multi-contact character movements. ACM Trans. Graph. 39, 4, Article 54 (July 2020), 14 pages.*
- *Sebastian Starke, Yiwei Zhao, Fabio Zinno, and Taku Komura. 2021. Neural animation layering for synthesizing martial arts movements. ACM Trans. Graph. 40, 4, Article 92 (August 2021), 16 pages.*

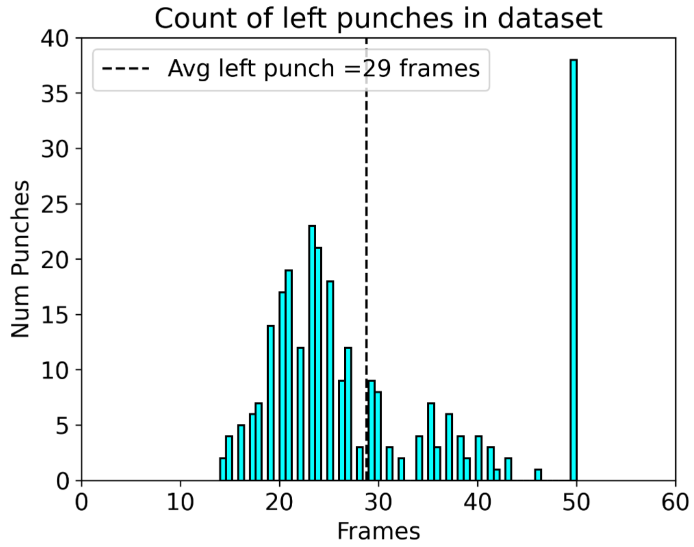
Thank You

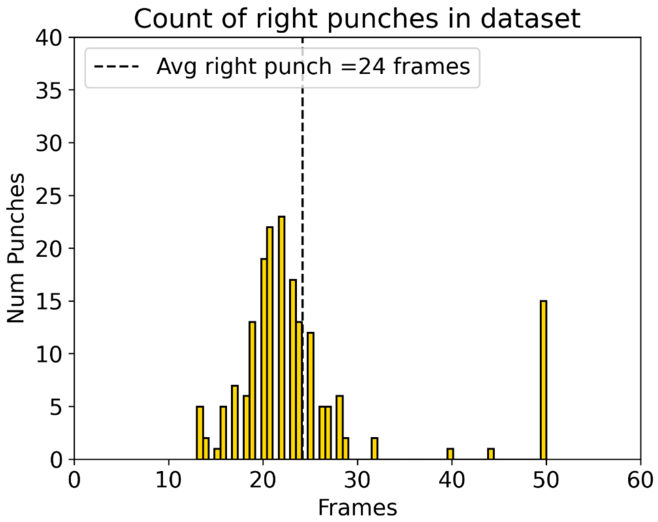
Supplimentary slides

Data Acquisition:
Punch annotation limitation

- Finding search threshold hard

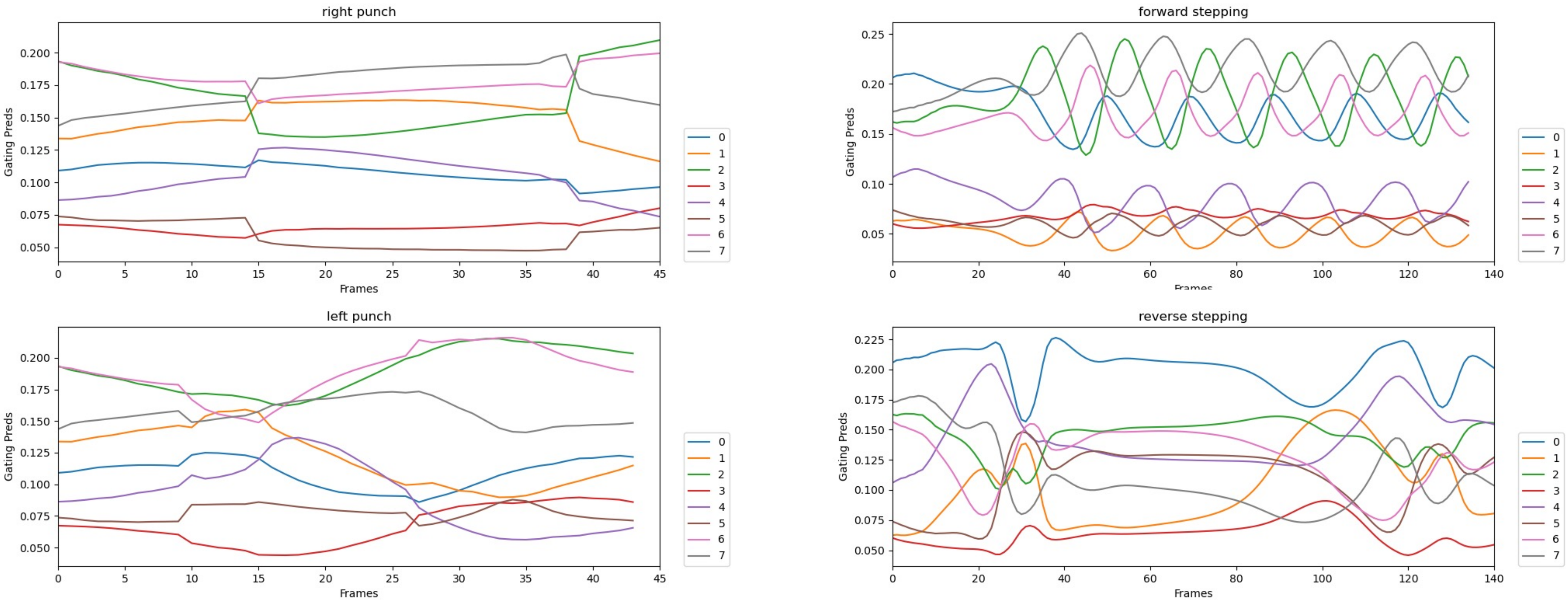






- Drop or keep slow punches?

Boxing neural model: Gating network output plots



Boxing controller system: Root Transformation

- Local positions $\mathrm{lp_i} = \mathrm{r_f} * (\mathrm{gp_i} - \mathrm{gp_0}|_{\mathrm{y}=0})$
- Local velocities and directions $\mathrm{l_i} = \mathrm{r_f} * \mathrm{g_i}$
- Rotations:
  - Compute forward facing directions
    - Vector between shoulder joints
    - Upward direction vector
    - Cross product gives forward directions
  - Rotations between user specified dir and fwd dir

Boxing controller system: Merge User input

- Trajectory root position

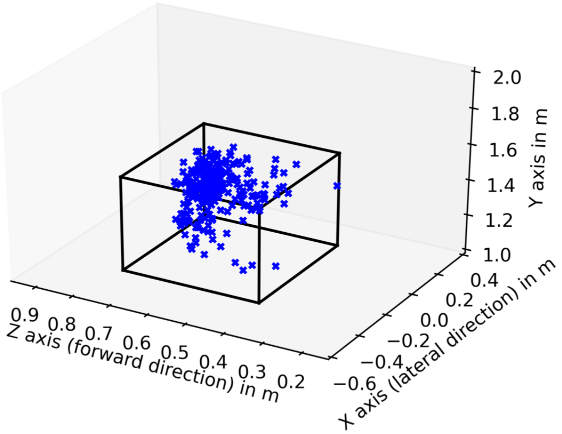
$$trp_i = trp_{i-1} + \left(1 - \left(\frac{(i - \frac{w}{2})}{\frac{w}{2}}\right)^{\tau}\right) \times (trp_i - trp_{i-1}) + \left(\frac{(i - \frac{w}{2})}{\frac{w}{2}}\right)^{\tau} \times v_{target}$$

- Trajectory wrist position

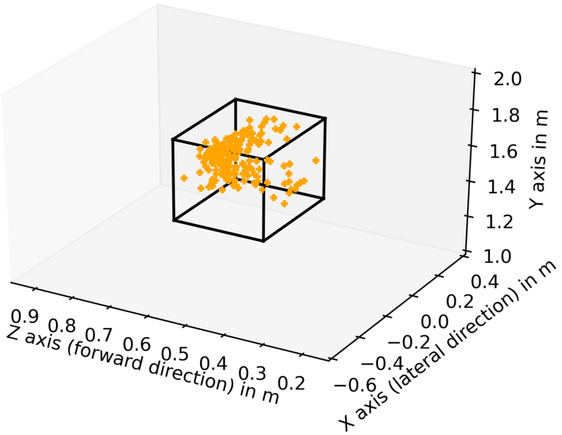
$$thp_i = thp_{i-1} + ((pt - hp_i) \times st) \times i$$

Punch control: Punch targets

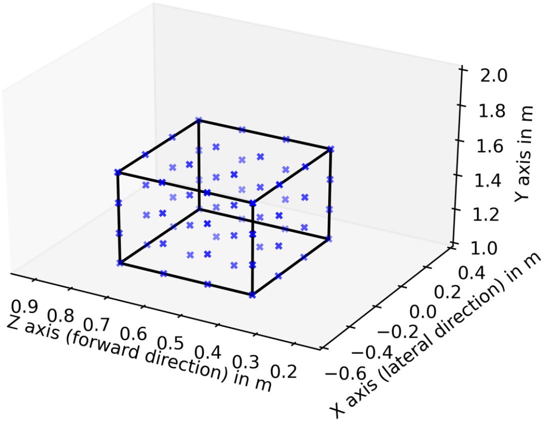
Dataset left punch targets



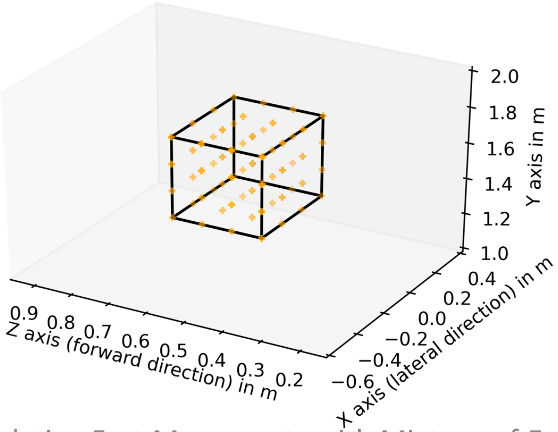
Dataset right punch targets



Uniform left punch targets

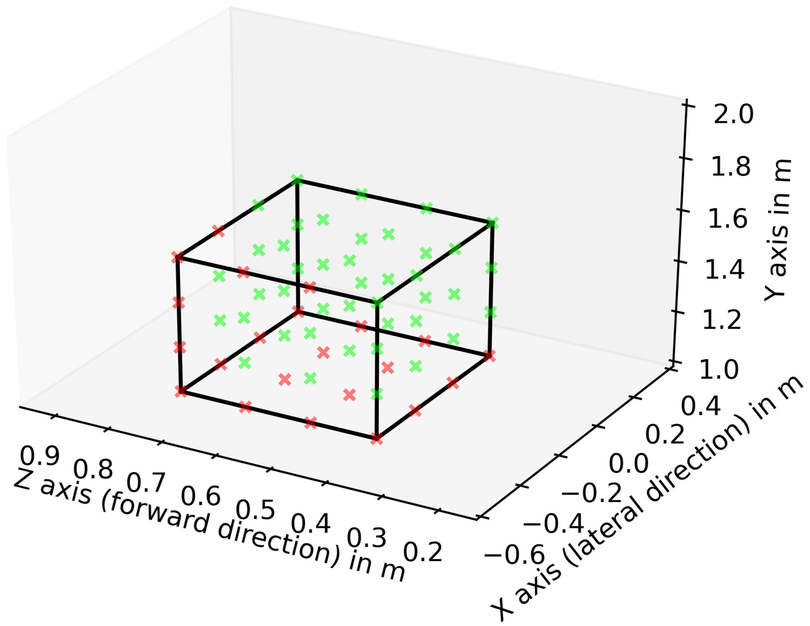


Uniform right punch targets

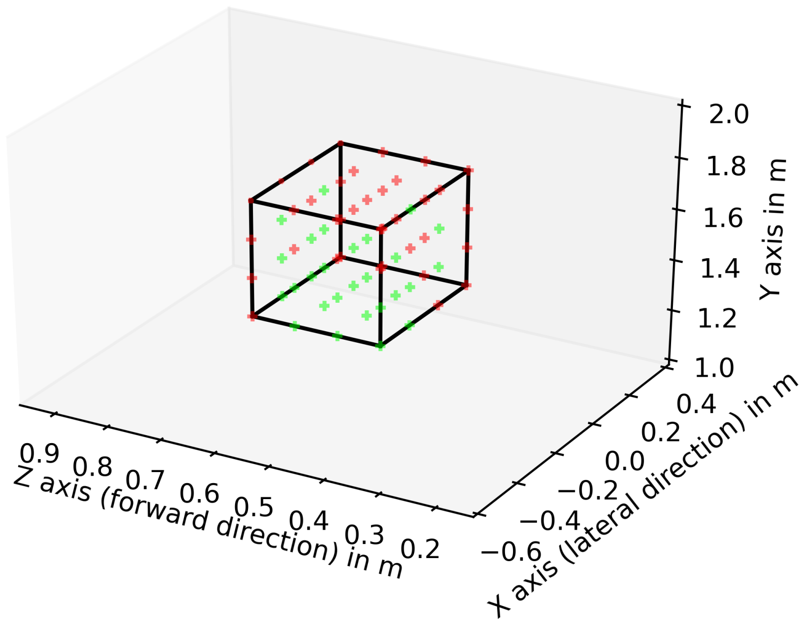


Punch control:
Punch accuracy plot (Traj Sampling)

Uniform left punch targets
with accuracy



Uniform right punch targets
with accuracy

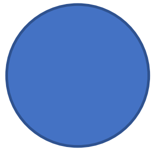


Punch control
Evaluation Metrics

- Punch Accuracy: How many punches are good enough?

$$accuracy = \begin{cases} 1; & \mathbf{l^2}(x,y) < p_{threshold} \\ 0; & \mathbf{l^2}(x,y) \geq p_{threshold} \end{cases}$$

$$\mathrm{PA} = \frac{B}{Z} \times 100$$



Avergae German male head

- Punch Error: How close do we get to target?

$$\mathrm{PE}(\mathbf{W}, \mathbf{P}) = \frac{1}{B} \sum_{n=1}^{B} \mathrm{MSE}(\mathbf{w}_n, \mathbf{p}_n)$$

- Foot Skating Error: Does the character float?

$$\mathrm{FSE} = \frac{1}{2} \sum_{l=1}^{2} d_l \left(2 - 2^{\left(\frac{h_l}{H}\right)}\right)$$

Punch control
Trajectory Sampling

- **Assumption:**
  - Traj similar to dataset punch
    ➔ more accurate punches
  - Models with trajectory window
    - covering approx. 27 frames (0.45s)
- **Observation:**
  - w=14, fs=3 leads to best accuracy.
  - Best model has somewhat high FSE.
  - Accuracy of left hand always higher than right hand.
- **Conclusion:**

  Sample according to dataset punches

w $(frames)$	fs $(frames)$	$PE_{overall}$ (cm)	$PA_{overall}$ (%)	FSE (cm)
4	1	11.43	20.31	**0.285**
6	1	11.29	28.91	0.305
8	1	11.02	28.91	0332
10	1	11.32	32.81	0.336
12	1	11.12	34.38	0.403
14	1	11.08	37.50	0.341
16	1	11.04	26.56	0.418
18	1	10.92	30.47	0.378
20	1	11.30	28.91	0.404
12	2	10.95	37.50	0.321
14	2	**10.20**	36.72	0.320
16	2	10.22	39.06	0.335
18	2	10.85	42.97	0.350
20	2	11.07	33.59	0.804
10	3	11.00	45.31	0.549
14	**3**	11.38	**53.91**	0.667
16	3	11.25	40.62	0.777
18	3	11.12	42.97	0.846

w $(frames)$	fs $(frames)$	$PE_{overall}$ (cm^2)	$PA_{overall}$ (%)	FSE (cm)
4	1	44.83	20.31	**0.285**
6	1	43.65	28.91	0.305
8	1	41.24	28.91	0332
10	1	43.79	32.81	0.336
12	1	42.19	34.38	0.403
14	1	42.06	37.50	0.341
16	1	41.51	26.56	0.418
18	1	40.57	30.47	0.378
20	1	43.50	28.91	0.404
12	2	40.81	37.50	0.321
14	2	**35.03**	36.72	0.320
16	2	35.26	39.06	0.335
18	2	40.01	42.97	0.350
20	2	41.47	33.59	0.804
10	3	41.43	45.31	0.549
14	**3**	44.19	**53.91**	0.667
16	3	43.16	40.62	0.777
18	3	42.29	42.97	0.846

Punch control Ablation Study

❑ Assumption:

- Sufficient trajectory information needed ➔ accurate punches
- 3 models using w=14 and fs=3

❑ Observation:

- Fewer trajectory variables leads to more steps
- All trajectory variables correctly move the hand to punch

❑ Conclusion:

Traj pos + vel ➔ accurate punches

Wrist trajectory variables	$PE_{overall}$ (cm)	$PA_{overall}$ $(\%)$	FSE (cm)
none	**11.18**	31.25	1.177
pos	12.13	35.94	0.980
pos + vel	11.38	**53.91**	**0.667**

Wrist trajectory variables	$PE_{overall}$ (cm^2)	$PA_{overall}$ $(\%)$	FSE (cm)
none	44.77	31.25	1.177
pos	50.02	35.94	0.980
pos + vel	**44.19**	**53.91**	**0.667**

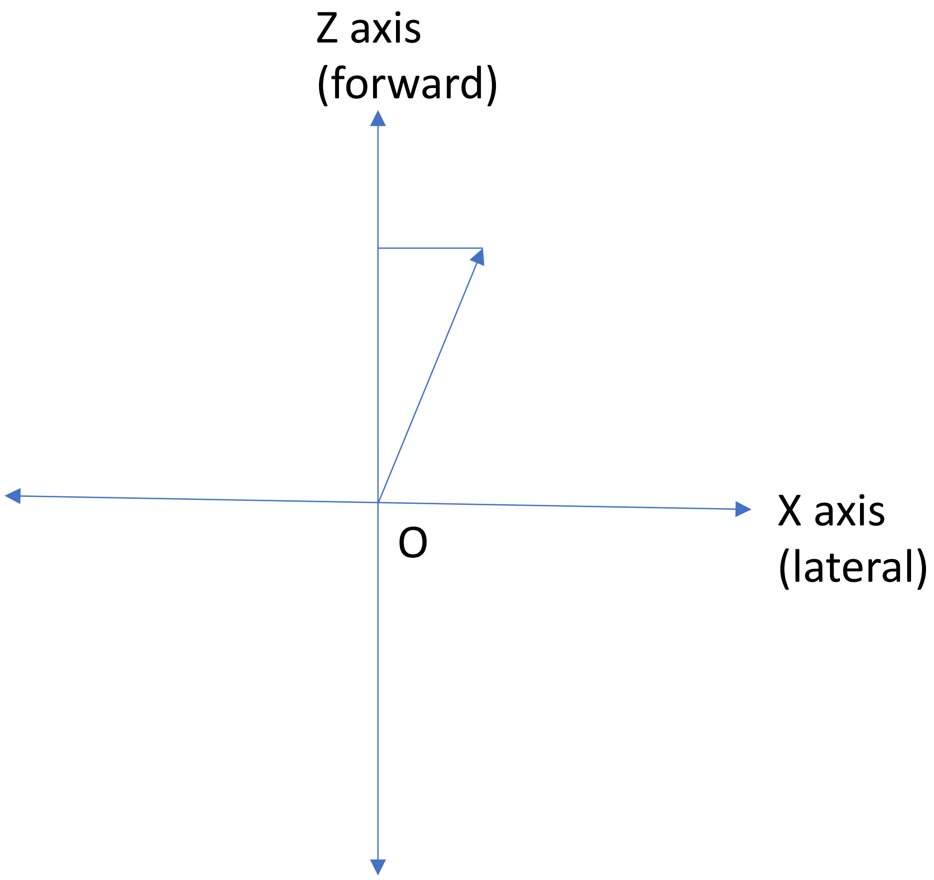
Stepping control
Evaluation Metrics

- ❑ Stepping Path Error (SPE): Path following capability

$$\mathrm{SPE}(\mathbf{G}, \mathbf{E}) = \frac{1}{N} \sum_{f=1}^{N} \mathrm{MSE}(\mathbf{g_f}, \mathbf{e_f})$$

- ❑ Foot Skating Error (FSE): Does the character float?

$$\mathrm{FSE} = \frac{1}{2} \sum_{l=1}^{2} d_l \left(2 - 2^{\left(\frac{h_l}{H}\right)}\right)$$



Stepping control
Trajectory Sampling

❑ Assumption:

- Trajectory similar to average stepping duration
  - ➔ better stepping
- Trained models: 0.3s to 2s windows
- Dataset average: 20 frames (0.3s)

w $(frames)$	fs $(frames)$	$SPE_{overall}$ (cm)	FSE (cm)
10	**2**	**33.61**	**0.52**
12	2	37.36	0.70
14	2	49.75	0.76
16	2	68.92	0.56
18	2	65.23	0.97
10	3	37.84	1.28
8	4	63.42	1.36
14	4	52.14	2.59

w $(frames)$	fs $(frames)$	$SPE_{overall}$ (cm^2)	FSE (cm)
10	**2**	**746.29**	**0.52**
12	2	1059.91	0.70
14	2	2773.14	0.76
16	2	3589.26	0.56
18	2	2959.33	0.97
10	3	1234.06	1.28
8	4	4724.17	1.36
14	4	3129.29	2.59

❑ Observation:

- Forward stepping better than backward stepping

❑ Conclusion:

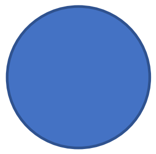
Sample according to dataset stepping action

Punch control
Evaluation Metrics

- Punch Accuracy: How many punches are good enough?

$$accuracy = \begin{cases} 1; & \mathbf{l^2}(x, y) < p_{threshold} \\ 0; & \mathbf{l^2}(x, y) \geq p_{threshold} \end{cases}$$

$$\text{PA} = \frac{B}{Z} \times 100$$



Avergae German male head

- Punch Error: How close do we get to target?

$$\text{PE}(\mathbf{W}, \mathbf{P}) = \frac{1}{B} \sum_{n=1}^{B} \mathbf{l^2}(\mathbf{w}_n, \mathbf{p}_n)$$

- Foot Skating Error: Does the character float?

$$\text{FSE} = \frac{1}{2} \sum_{l=1}^{2} d_l \left(2 - 2^{\left(\frac{h_l}{H}\right)} \right)$$

Punch control
Trajectory Sampling

- Assumption:
  - Traj similar to dataset punch
    - ➔ more accurate punches
  - Models with trajectory window
    - covering approx. 27 frames (0.45s)
- Observation:
  - w=14, fs=3 leads to best accuracy.
  - Best model has somewhat high FSE.
  - Accuracy of left hand always higher than right hand.
- Conclusion:

  Sample according to dataset punches

w $(frames)$	fs $(frames)$	$PE_{overall}$ (cm)	$PA_{overall}$ (%)	FSE (cm)
4	1	11.43	20.31	**0.285**
6	1	11.29	28.91	0.305
8	1	11.02	28.91	0332
10	1	11.32	32.81	0.336
12	1	11.12	34.38	0.403
14	1	11.08	37.50	0.341
16	1	11.04	26.56	0.418
18	1	10.92	30.47	0.378
20	1	11.30	28.91	0.404
12	2	10.95	37.50	0.321
14	2	**10.20**	36.72	0.320
16	2	10.22	39.06	0.335
18	2	10.85	42.97	0.350
20	2	11.07	33.59	0.804
10	3	11.00	45.31	0.549
14	**3**	11.38	**53.91**	0.667
16	3	11.25	40.62	0.777
18	3	11.12	42.97	0.846

Punch control
Ablation Study

- ❑ Assumption:
  - Sufficient trajectory information needed
  - 3 models using w=14 and fs=3
- ❑ Observation:
  - Fewer trajectory variables leads to more steps
  - All trajectory variables correctly move the hand to punch

Wrist trajectory variables	$PE_{overall}$ (cm)	$PA_{overall}$ $(\%)$	FSE (cm)
none	**11.18**	31.25	1.177
pos	12.13	35.94	0.980
pos + vel	11.38	**53.91**	**0.667**

- ❑ Conclusion:

  Traj pos + vel ➔ accurate punches

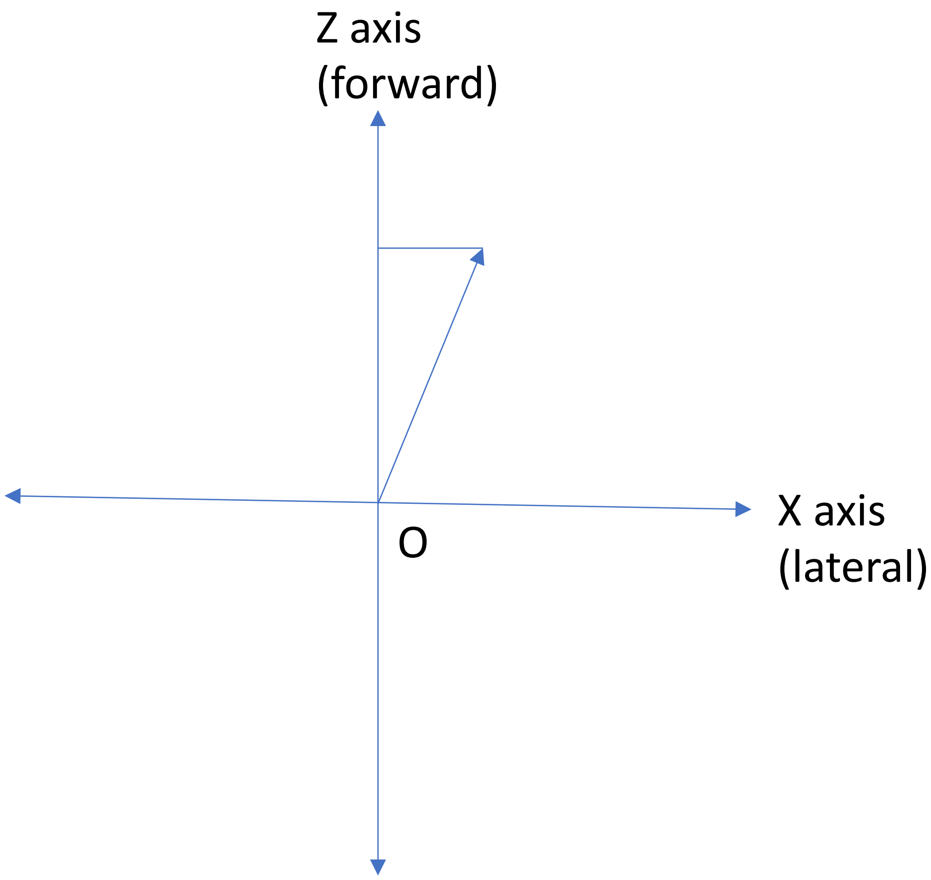
Stepping control
Evaluation Metrics

- Stepping Path Error (SPE): Path following capability

$$\mathrm{SPE}(\mathbf{G}, \mathbf{E}) = \frac{1}{N} \sum_{f=1}^{N} \mathbf{l}^{\mathbf{2}}(\mathbf{g_f}, \mathbf{e_f})$$

- Foot Skating Error (FSE): Does the character float?

$$\mathrm{FSE} = \frac{1}{2} \sum_{l=1}^{2} d_l \left(2 - 2^{\left(\frac{h_l}{H}\right)}\right)$$



Stepping control
Trajectory Sampling

❑ Assumption:

- Trajectory window similar to average stepping duration
  ➔ better stepping
- Trained models: 0.3s to 2s windows
- Dataset average: 20 frames (0.3s)

w $(frames)$	fs $(frames)$	$SPE_{overall}$ (cm)	FSE (cm)
10	**2**	**33.61**	**0.52**
12	2	37.36	0.70
14	2	49.75	0.76
16	2	68.92	0.56
18	2	65.23	0.97
10	3	37.84	1.28
8	4	63.42	1.36
14	4	52.14	2.59

❑ Observation:

- Forward stepping better than backward stepping

❑ Conclusion:

Sample according to dataset stepping action