Render the Possibilities

SIGGRAPH2016

THE **43RD** INTERNATIONAL
CONFERENCE AND **EXHIBITION** ON

& Computer Graphics Interactive Techniques

24-28 JULY

ANAHEIM, CALIFORNIA

Render the Possibilities
SIGGRAPH2016

THE 43RD INTERNATIONAL
CONFERENCE AND EXHIBITION ON
Computer Graphics
& Interactive Techniques

A Deep Learning Framework for Character Motion Synthesis and Editing

Daniel Holden *, Jun Saito †, Taku Komura *,

*The University of Edinburgh
†Marza Animation Planet

Outline

Motivation

Synthesis

Editing

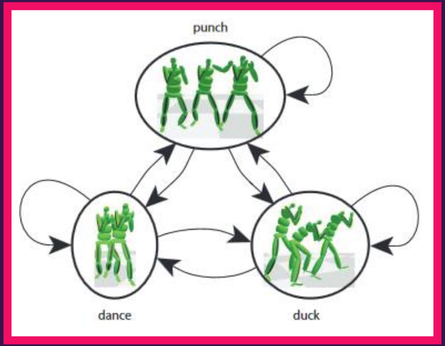
Discussion

Goal

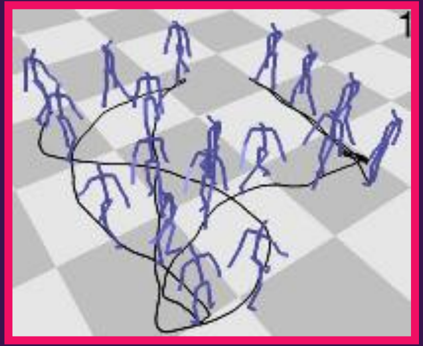
Data driven synthesis of motion from high level controls with no manual preprocessing

Previous Work

- Lots of manual processing (Graphs, Trees)
  - Segmentation
  - Alignment
  - Classification



[Heck et al. 2007]

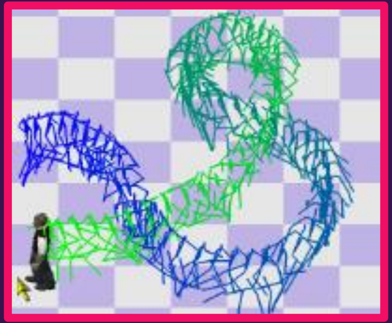


[Kovar et al. 2002]

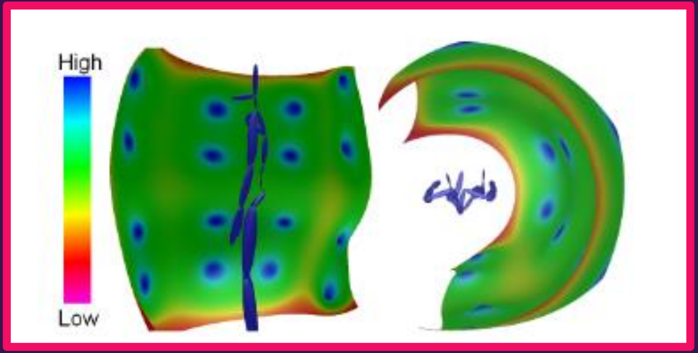
Previous Work

- Scalability Issues (RBF, GP, GPLVM, kNN)
  - Must store whole database in memory
  - Grows $O(n^2)$ with number of data points
  - Requires expensive acceleration structures

[Lee et al. 2010]

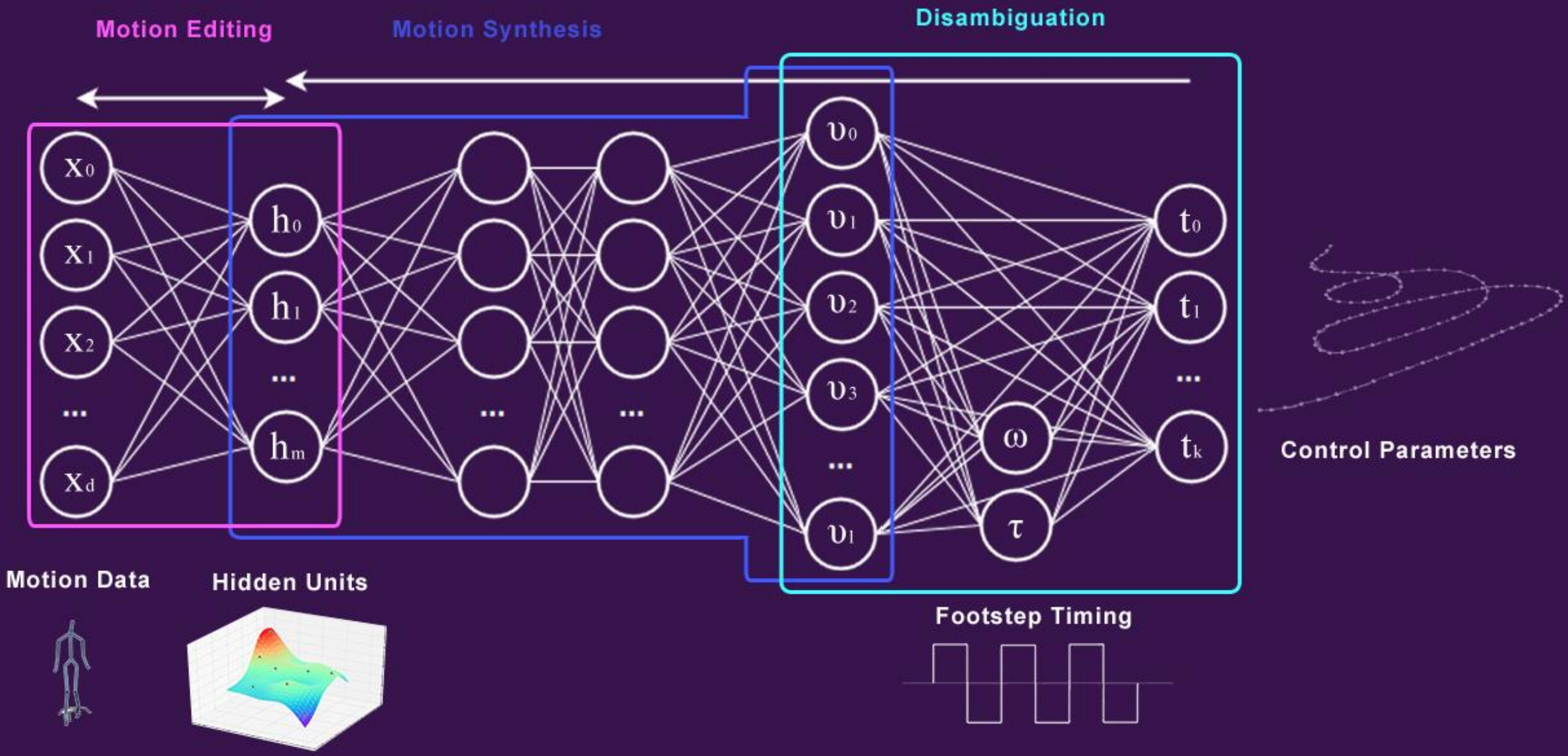


[Park et al. 2002]

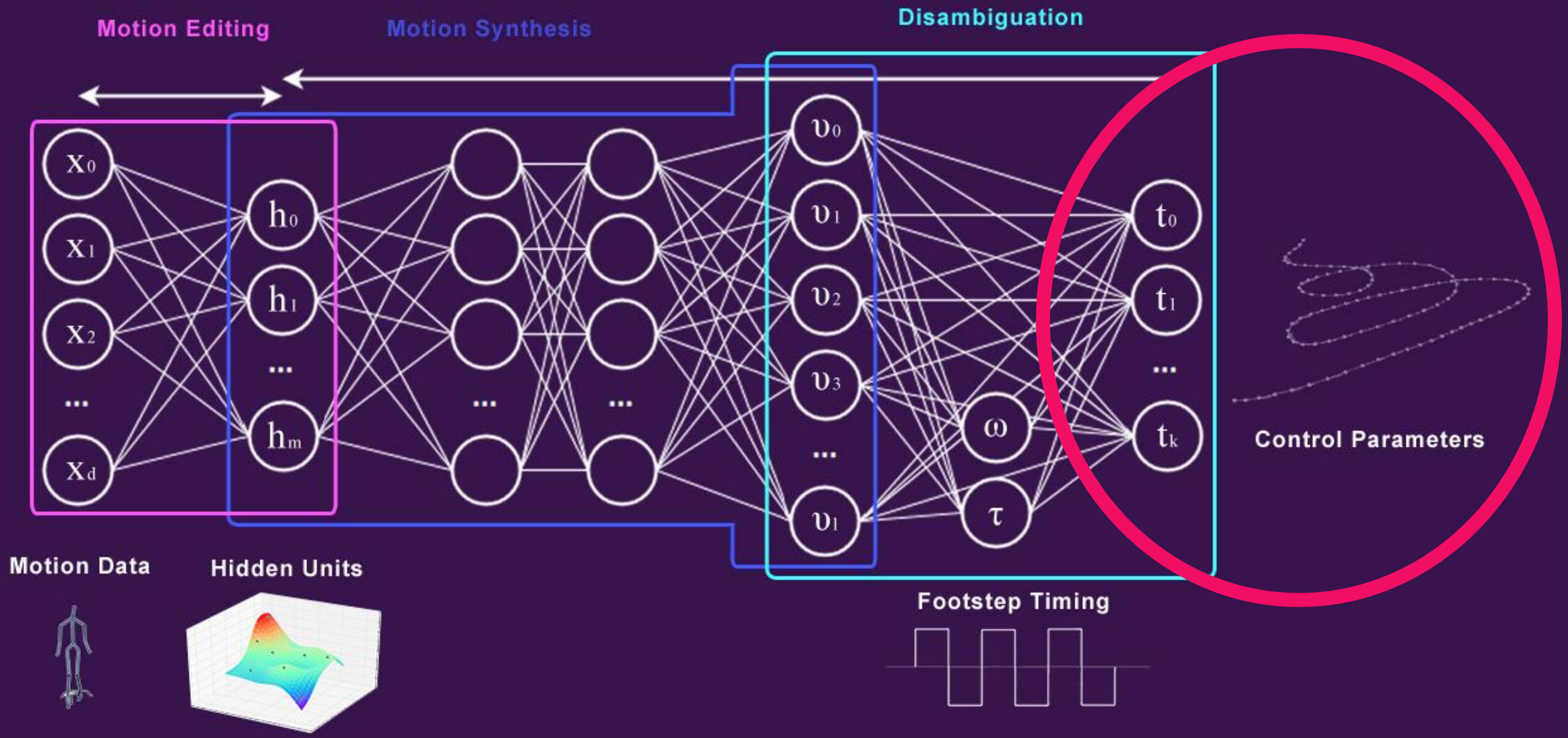


[Mukai and Kuriyama 2005]

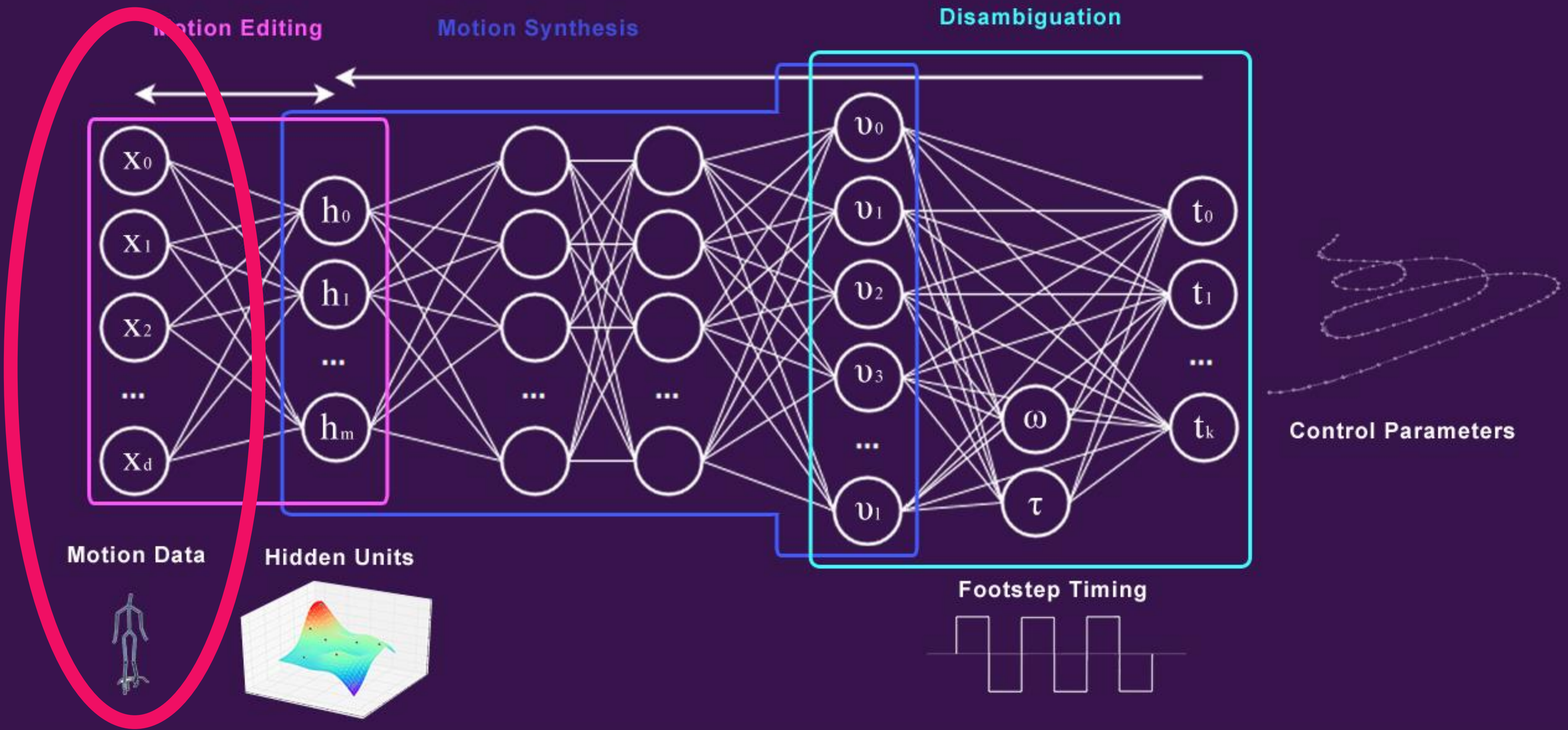
Overview


Motion Editing
Motion Synthesis
Disambiguation
x0
x1
x2
...
xd
h0
h1
...
hm
...
...
υ0
υ1
υ2
υ3
...
υl
ω
τ
t0
t1
...
tk
Control Parameters
Motion Data
Hidden Units
Footstep Timing

Overview


Motion Editing
Motion Synthesis
Disambiguation
x0
x1
x2
...
xd
h0
h1
...
hm
...
...
υ0
υ1
υ2
υ3
...
υl
ω
τ
t0
t1
...
tk
Control Parameters
Motion Data
Hidden Units
Footstep Timing

Overview


Motion Editing
Motion Synthesis
Disambiguation
x_0
x_1
x_2
...
x_d
h_0
h_1
...
h_m
...
...
υ_0
υ_1
υ_2
υ_3
...
υ_l
ω
τ
t_0
t_1
...
t_k
Control Parameters
Motion Data
Hidden Units
Footstep Timing

Outline

Motivation

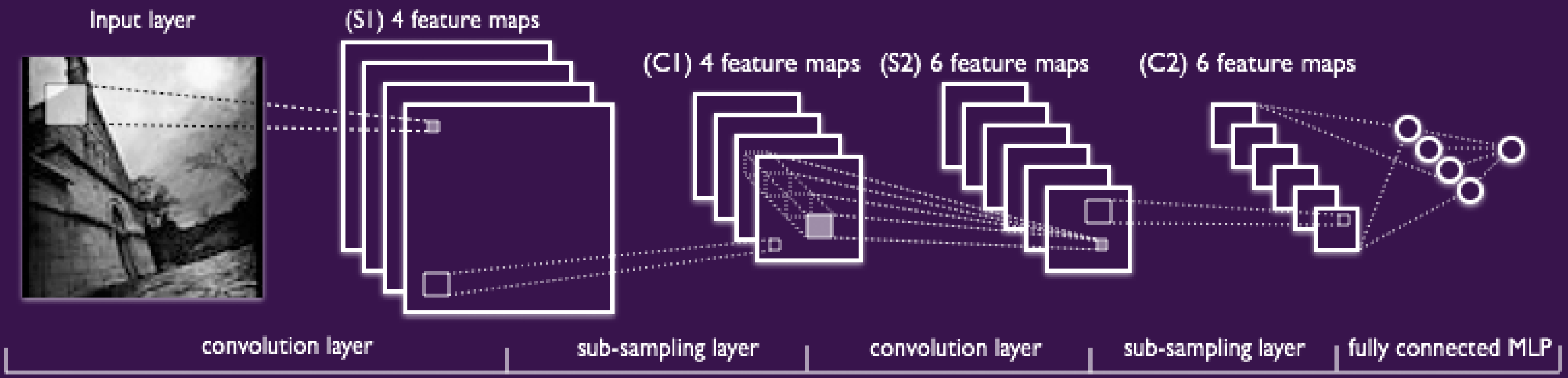
Synthesis

Editing

Discussion

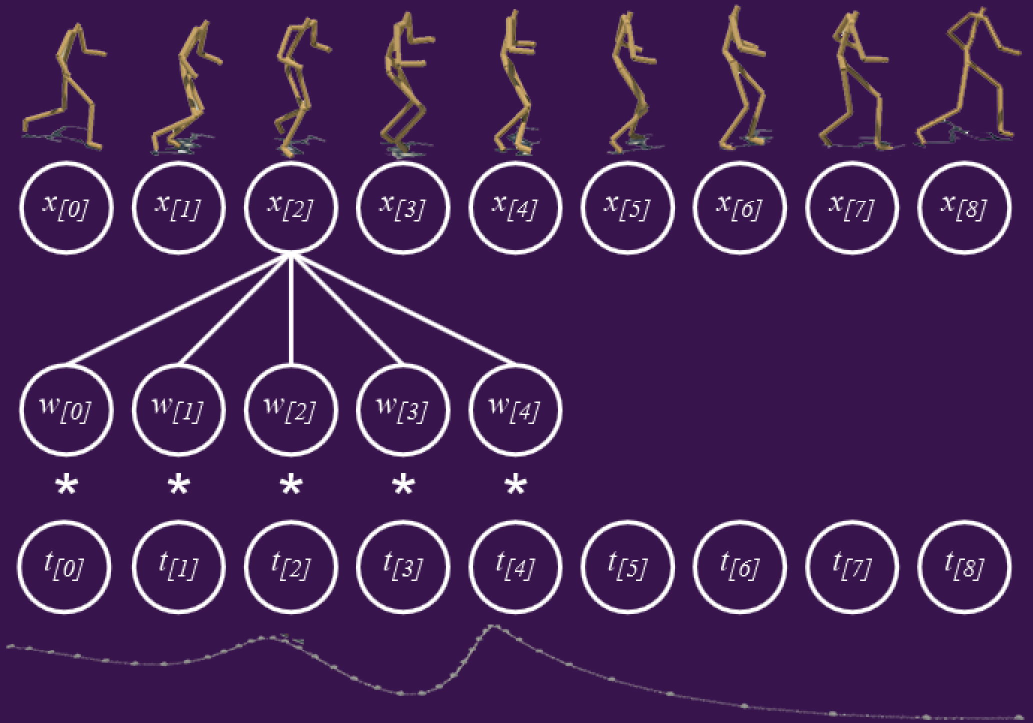
Convolutional Neural Networks

- Great success in classification and segmentation for images, video, sound
- We can use CNN on motion data too



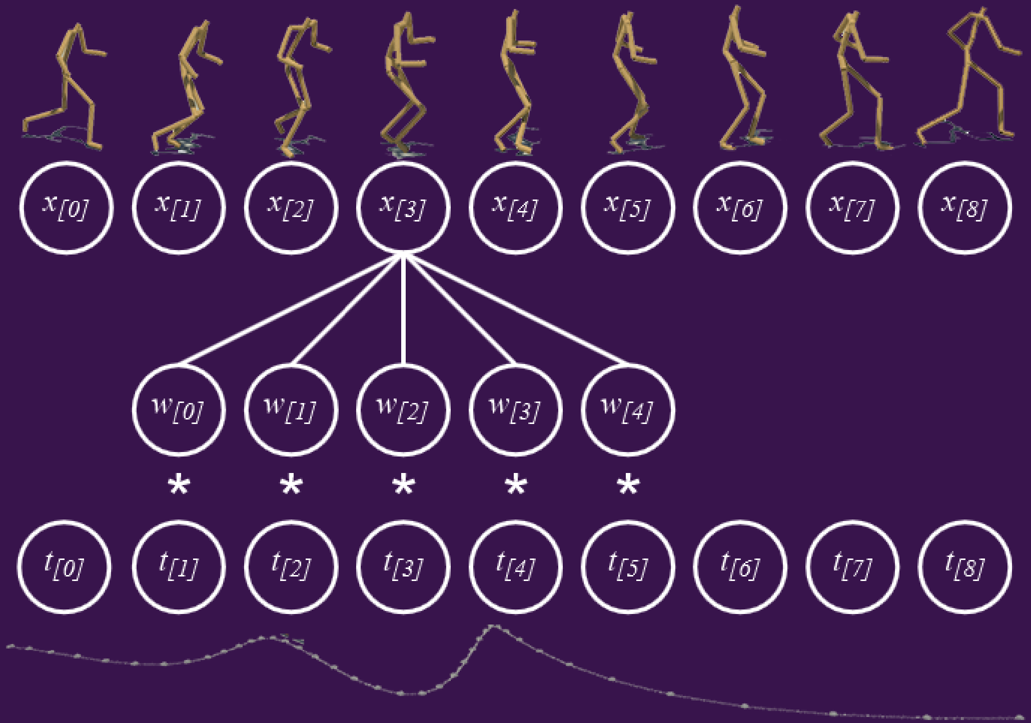
Convolution

Filters convolve over temporal dimension



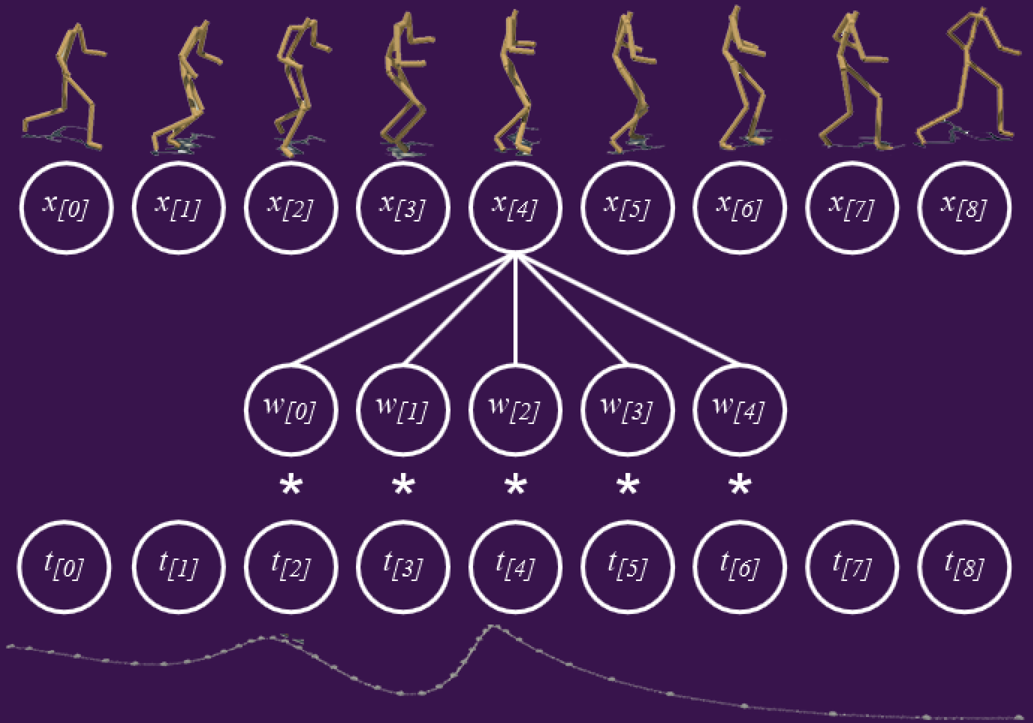
Convolution

Filters convolve over temporal dimension



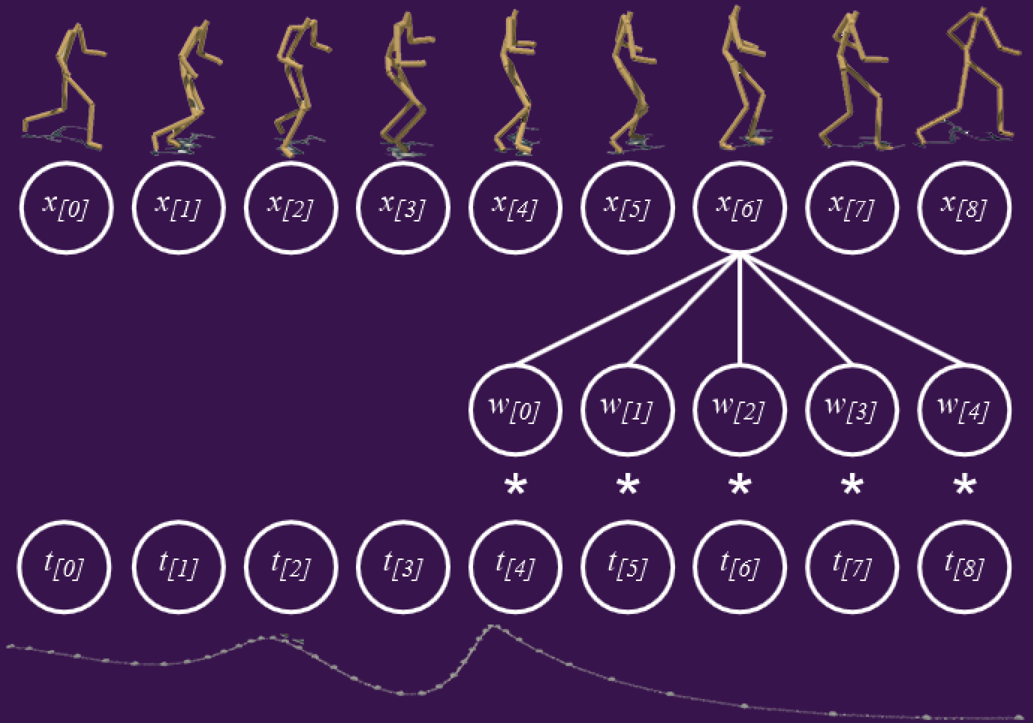
Convolution

Filters convolve over temporal dimension



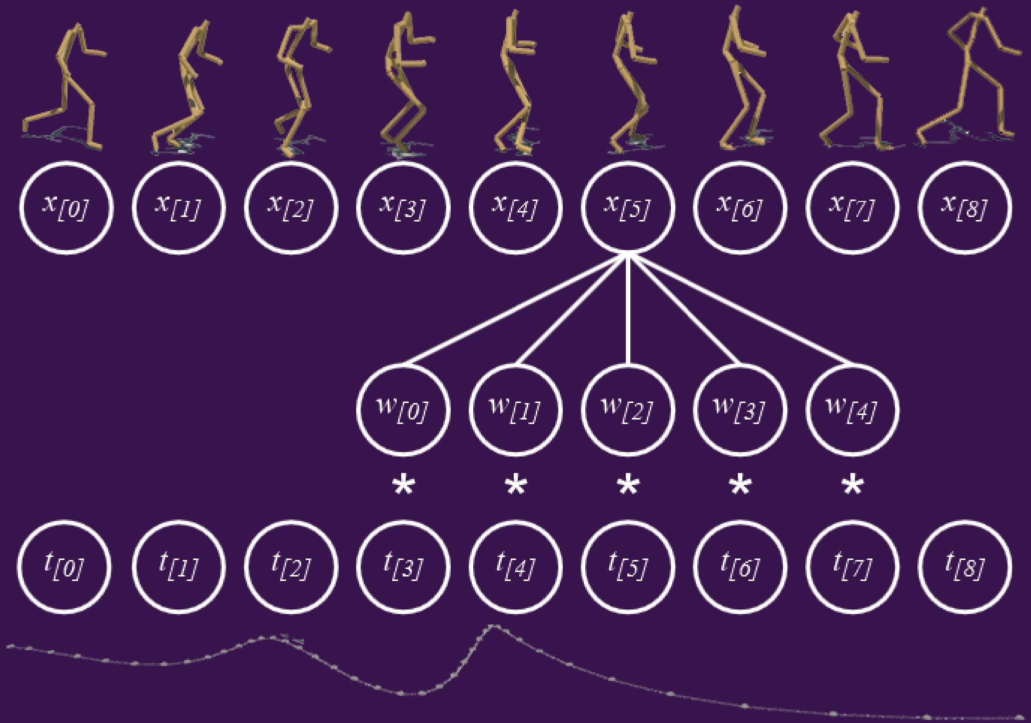
Convolution

Filters convolve over temporal dimension



Convolution

Filters convolve over temporal dimension



Outline

Motivation

Synthesis

Editing

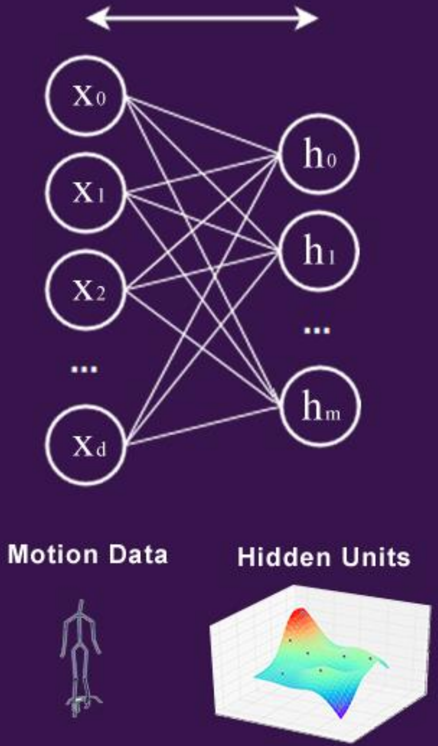
Discussion

Motion Editing

Post processing may not ensure naturalness

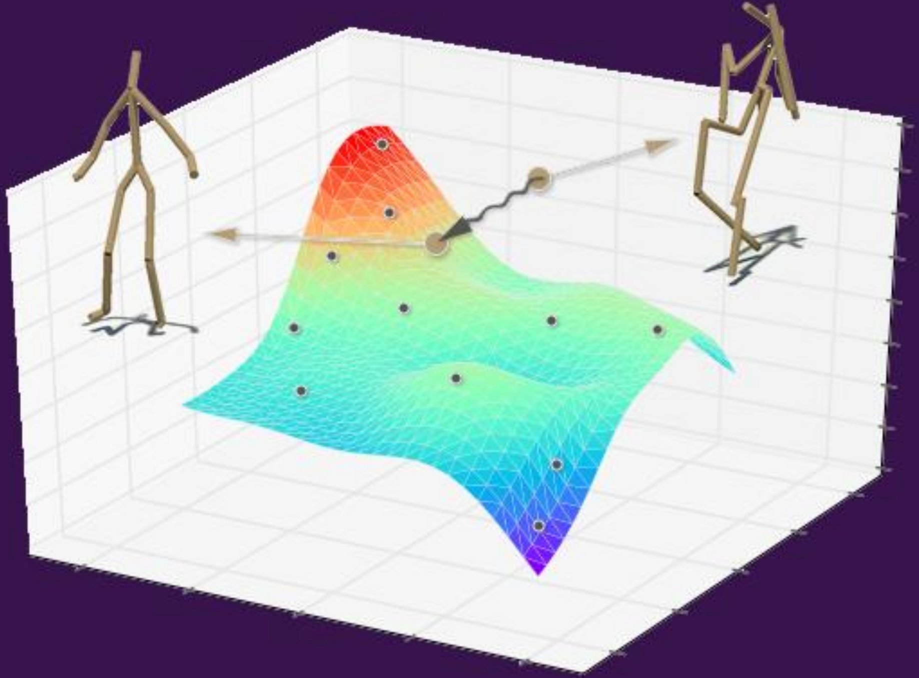
Motion Editing

- We edit using the motion manifold learned by a Convolutional Autoencoding Network [Holden et al. 2015]



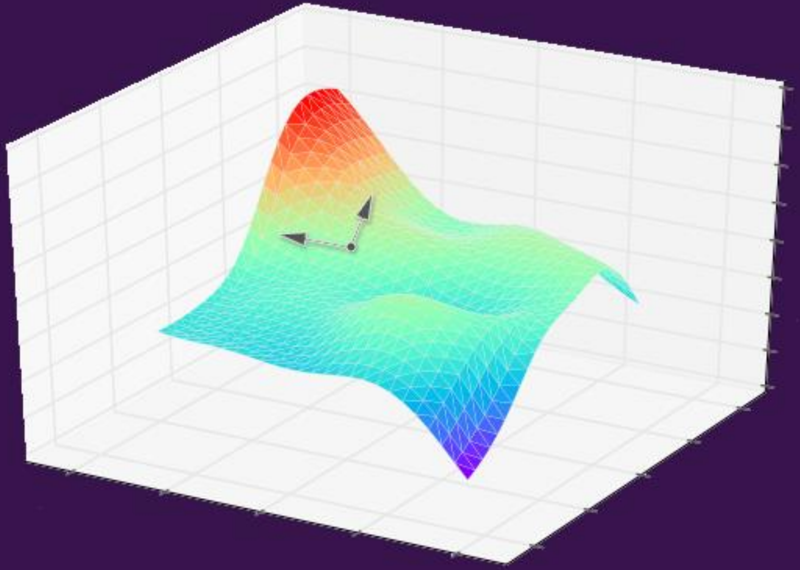
Autoencoder

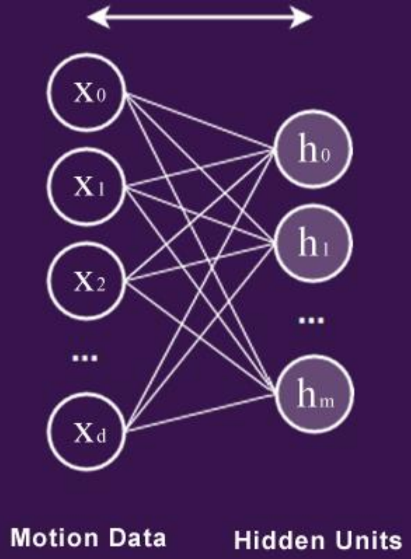
- Learns *projection operator* of motion manifold



Manifold Surface

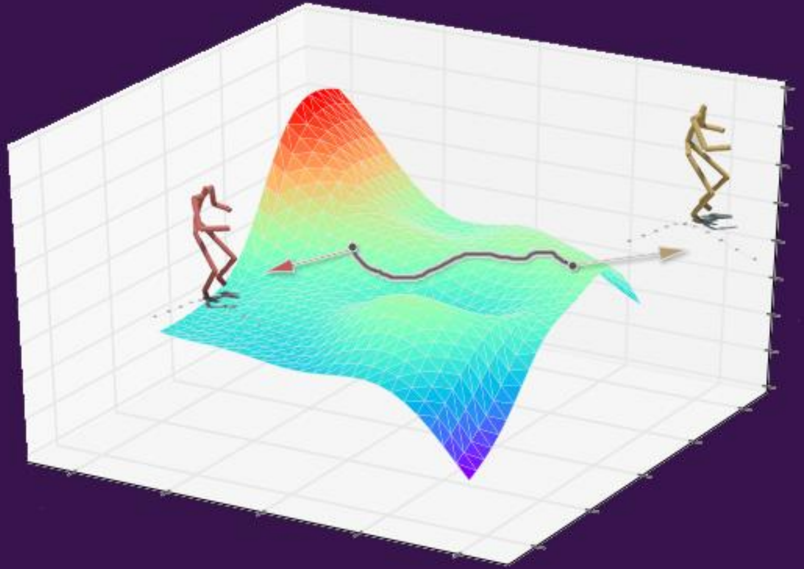
- *Hidden Unit* values parametrise manifold surface
- Adjusting them ensures motion remains natural

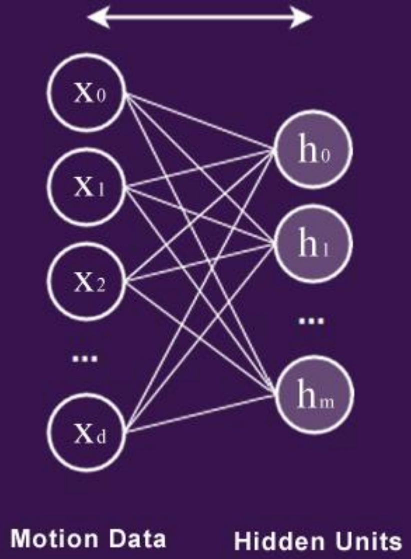




Constraint Satisfaction

- Motion editing is a *constraint satisfaction problem* over *Hidden Units*





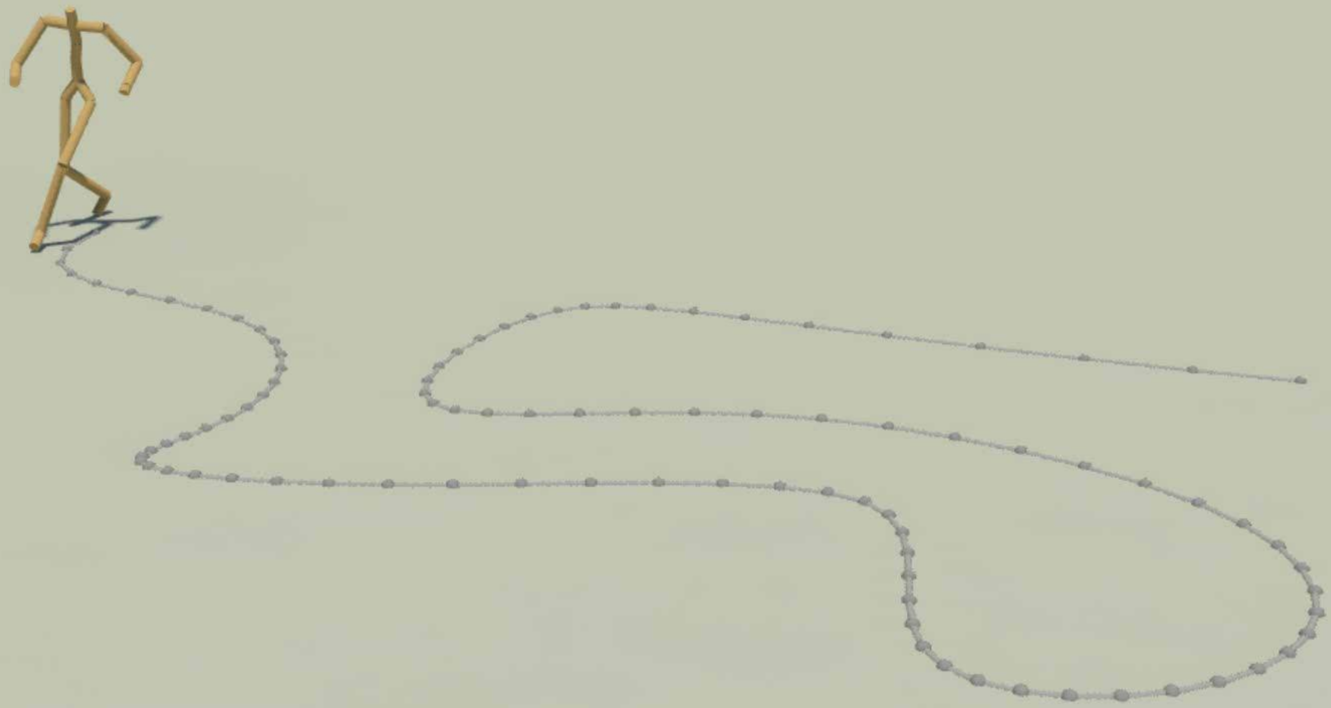
Constraint Satisfaction

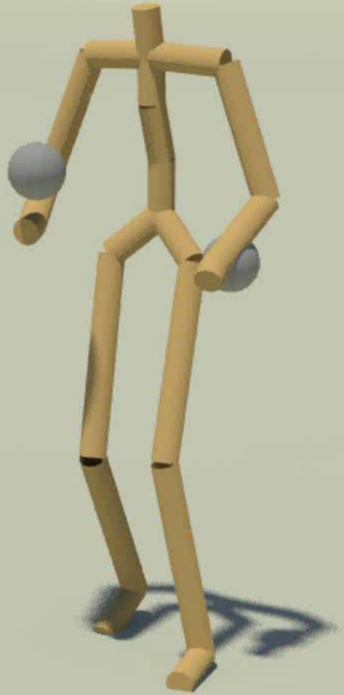
- Local foot velocity must equal global velocity

$$Pos(\mathrm{H}) = \sum_j \|\mathbf{v}_r^{\mathrm{H}} + \omega^{\mathrm{H}} \times \mathbf{p}_j^{\mathrm{H}} + \mathbf{v}_j^{\mathrm{H}} - \mathbf{v}'_j\|_2^2.$$

- Output trajectory must equal input trajectory

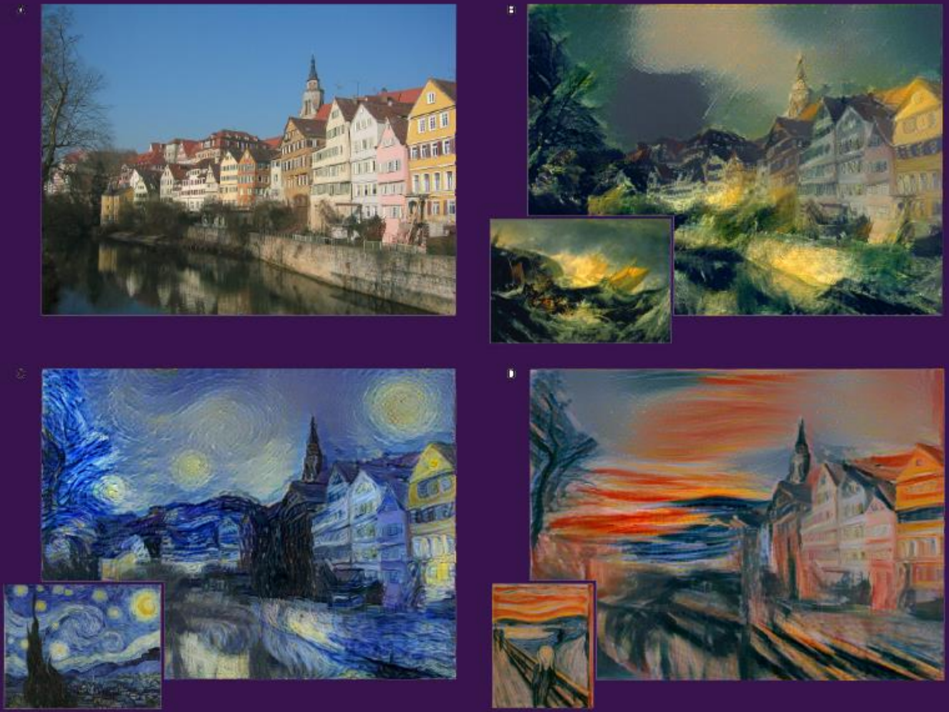
$$Traj(\mathrm{H}) \quad = \quad \|\omega^{\mathrm{H}} - \omega'\|_2^2 + \|\mathbf{v}_r^{\mathrm{H}} - \mathbf{v}'_r\|_2^2$$





A Neural Algorithm of Artistic Style

- Combine style of one image with content of another [Gatys et al. 2015]



Style Constraint

- Gram Matrix of *Hidden Units* encode style
- Actual Values of *Hidden Units* encode content

$$Style(\mathbf{H}) = s\|G(\Phi(\mathbf{S})) - G(\mathbf{H})\|_2^2 + c\|\Phi(\mathbf{C}) - \mathbf{H}\|_2^2$$

$$G(\mathbf{H}) = \frac{\sum_i^n \mathbf{H}_i \mathbf{H}_i^T}{n}$$

- **No correspondence between clips required!**

Style Constraint

- Gram Matrix of *Hidden Units* encode style
- Actual Values of *Hidden Units* encode content

$$Style(\mathbf{H}) = s\|G(\mathbf{\Phi}(\mathbf{S})) - G(\mathbf{H})\|_2^2 + \underbrace{c\|\mathbf{\Phi}(\mathbf{C}) - \mathbf{H}\|_2^2}_{\text{Content Term}}$$

$$G(\mathbf{H}) = \frac{\sum_i^n \mathbf{H}_i \mathbf{H}_i^T}{n}$$

- **No correspondence between clips required!**

Style Constraint

- Gram Matrix of *Hidden Units* encode style
- Actual Values of *Hidden Units* encode content

$$Style(\mathbf{H}) = s\|G(\Phi(\mathbf{S})) - G(\mathbf{H})\|_2^2 + c\|\Phi(\mathbf{C}) - \mathbf{H}\|_2^2$$

Style Term

$$G(\mathbf{H}) = \frac{\sum_i^n \mathbf{H}_i \mathbf{H}_i^T}{n}$$

- **No correspondence between clips required!**

Style Constraint

- Gram Matrix of *Hidden Units* encode style
- Actual Values of *Hidden Units* encode content

$$Style(\mathbf{H}) = s\|G(\Phi(\mathbf{S})) - G(\mathbf{H})\|_2^2 + c\|\Phi(\mathbf{C}) - \mathbf{H}\|_2^2$$

$$G(\mathbf{H}) = \frac{\sum_i^n \mathbf{H}_i \mathbf{H}_i^T}{n}$$

Gram Matrix

- **No correspondence between clips required!**

Style Constraint

- Gram Matrix of *Hidden Units* encode style
- Actual Values of *Hidden Units* encode content

$$Style(\mathbf{H}) = s\|G(\Phi(\mathbf{S})) - G(\mathbf{H})\|_2^2 + c\|\Phi(\mathbf{C}) - \mathbf{H}\|_2^2$$

$$G(\mathbf{H}) = \frac{\sum_i^n \mathbf{H}_i \mathbf{H}_i^T}{n}$$

- **No correspondence between clips required!**

Style

Content

Transfer

Outline

Motivation

Synthesis

Editing

Discussion

Training

- **Motion Manifold**
  - Several large databases (including whole CMU)
  - Training takes around 6 hours

- **Motion Synthesis**
  - Task specific data only (e.g. locomotion only)
  - Training takes around 1 hour

Contribution

- **High quality synthesis** with **no manual preprocessing**
- Motion synthesis and editing in **unified framework**
- **Procedural**, **parallel** technique

Future Work

- Need more general solution for ambiguity issue
- Wish to use more high level features with a deeper network
- What changes are required for interactive applications?