Learning to See and Generate People

Siyu Tang

Max Planck Institute for Intelligent Systems

Learning to see and generate people

Learning to see and generate **people**

Learning to see and generate people

- People are often a central element of visual scenes.

Learning to see and generate people

- People are often a central element of visual scenes.

Learning to see and generate people

- People are often a central element of visual scenes.

Learning to see and generate people

- People are often a central element of visual scenes.
- Visual understanding of people is the key component in many autonomous systems.

Learning to see and generate people

- People are often a central element of visual scenes.
- Visual understanding of people is the key component in many autonomous systems.

Autonomous driving

Learning to see and generate people

- People are often a central element of visual scenes.
- Visual understanding of people is the key component in many autonomous systems.

Autonomous driving

Mixed Reality

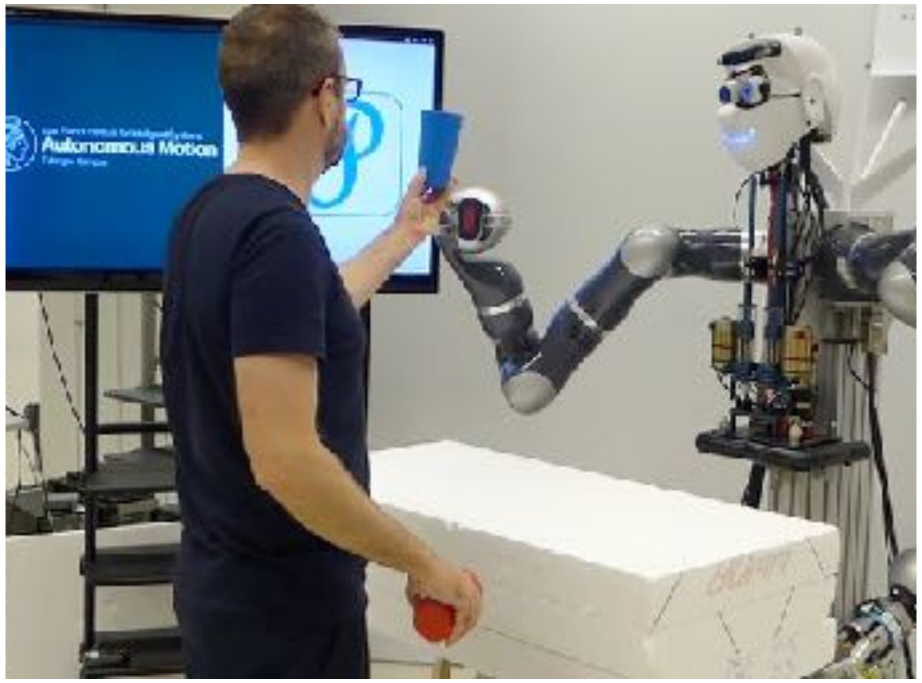
Learning to see and generate people

- People are often a central element of visual scenes.
- Visual understanding of people is the key component in many autonomous systems.

Autonomous driving

Mixed Reality

Human robot interaction



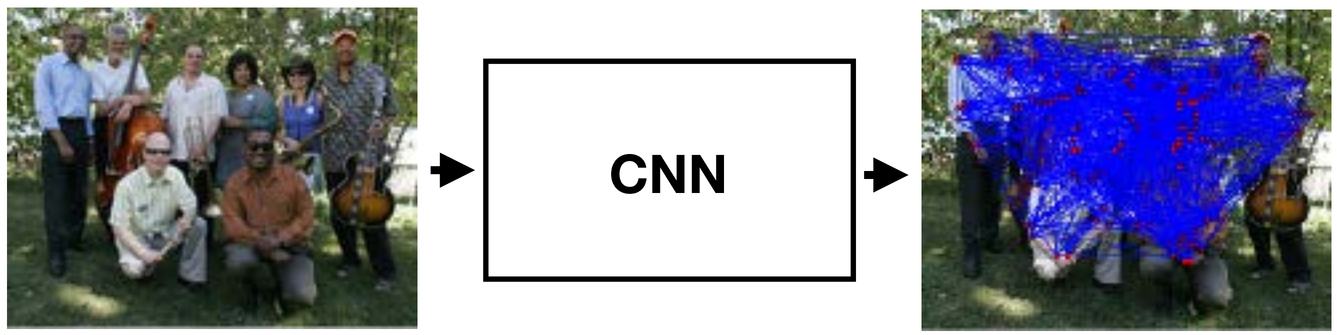
In this talk:

- Learning to see humans:

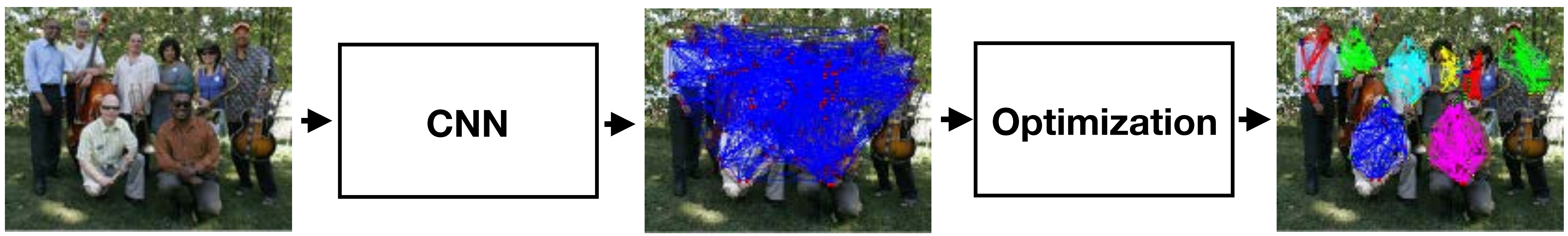
- Learning to generate humans:

Multi-person Pose Estimation by Graph Decomposition

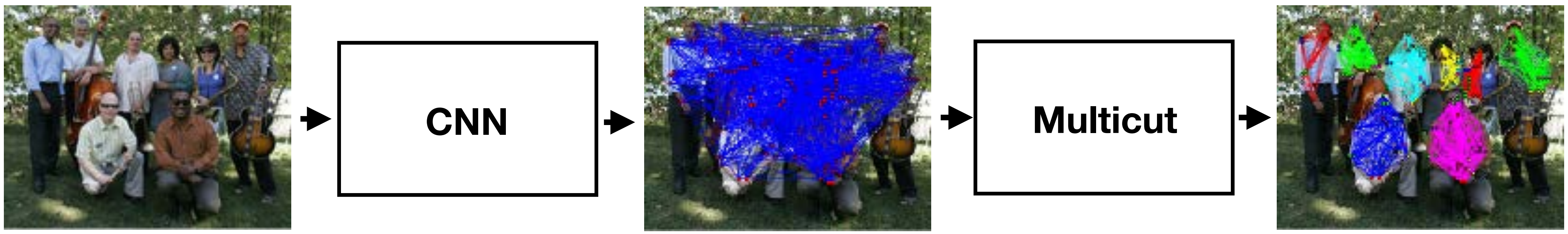
Multi-person Pose Estimation by Graph Decomposition



Multi-person Pose Estimation by Graph Decomposition

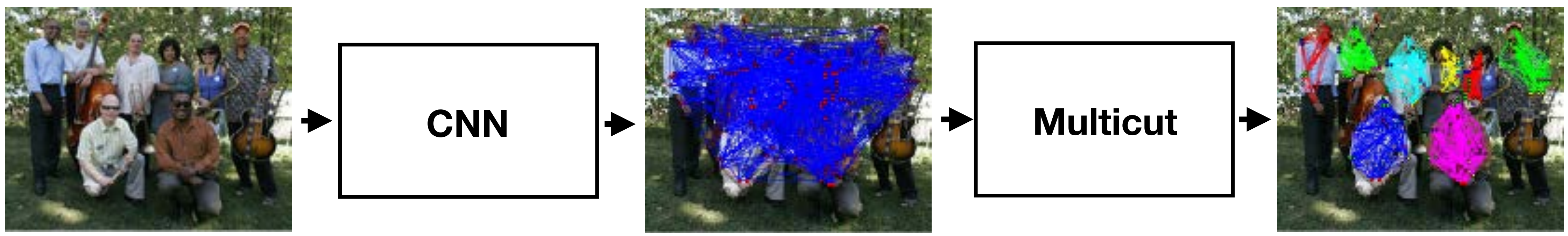


Multi-person Pose Estimation by Graph Decomposition



[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

Multi-person Pose Estimation by Graph Decomposition



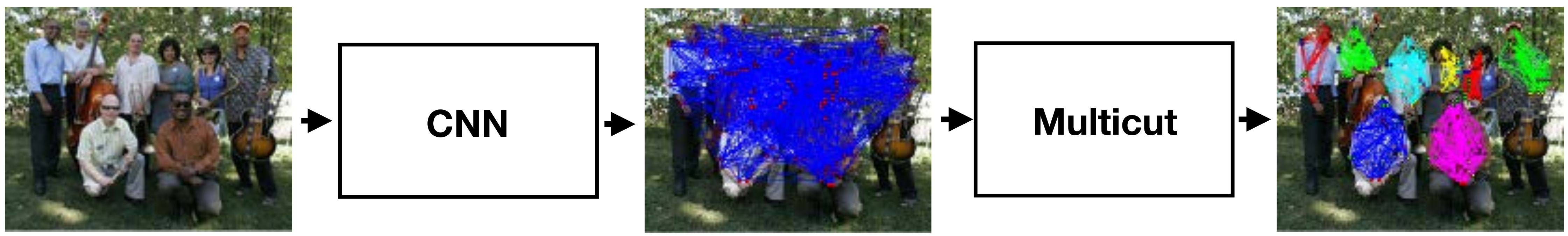
[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

- Minimum Cost Multicut Problem [Chopra et al. The partition problem. *Mathematical Programming 1993*]

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \mathrm{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad .$$

Multi-person Pose Estimation by Graph Decomposition

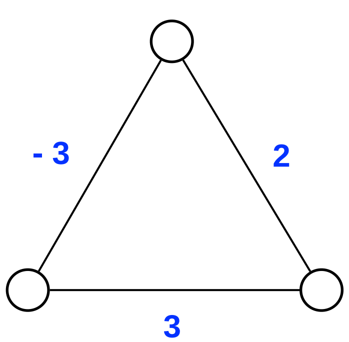


[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

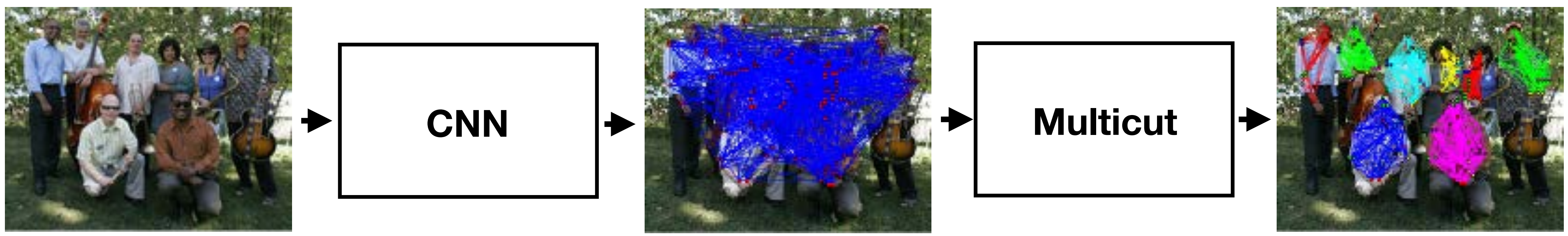
- Minimum Cost Multicut Problem [Chopra et al. The partition problem. *Mathematical Programming 1993*]

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \text{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad .$$



Multi-person Pose Estimation by Graph Decomposition

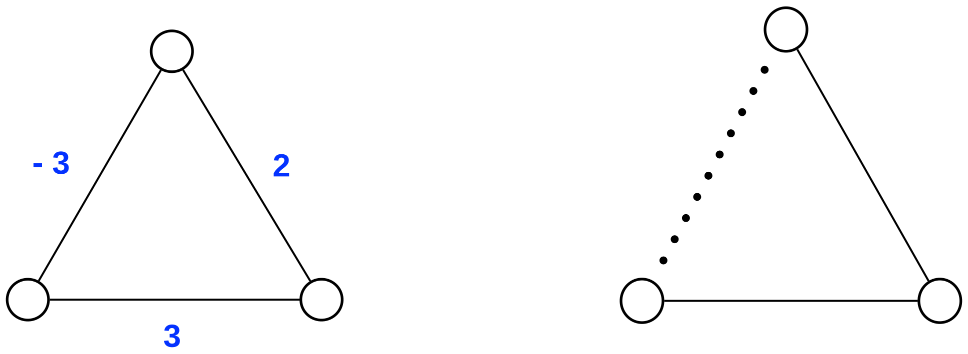


[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

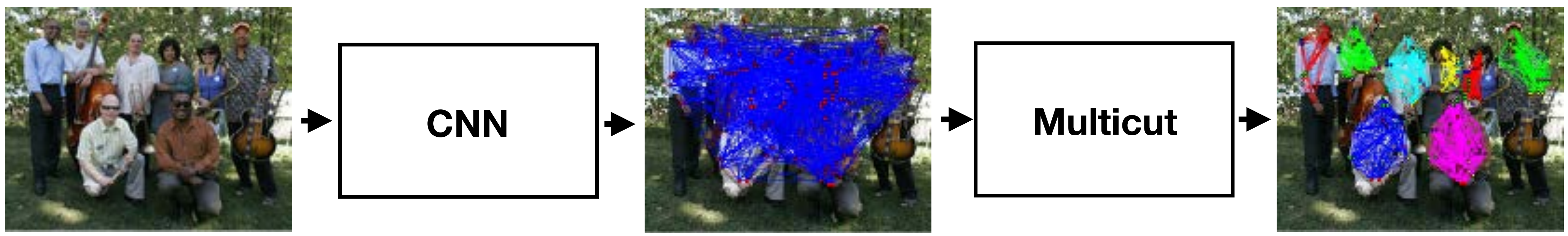
- Minimum Cost Multicut Problem [Chopra et al. The partition problem. *Mathematical Programming 1993*]

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \mathrm{cc}(G)\, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \ .$$



Multi-person Pose Estimation by Graph Decomposition

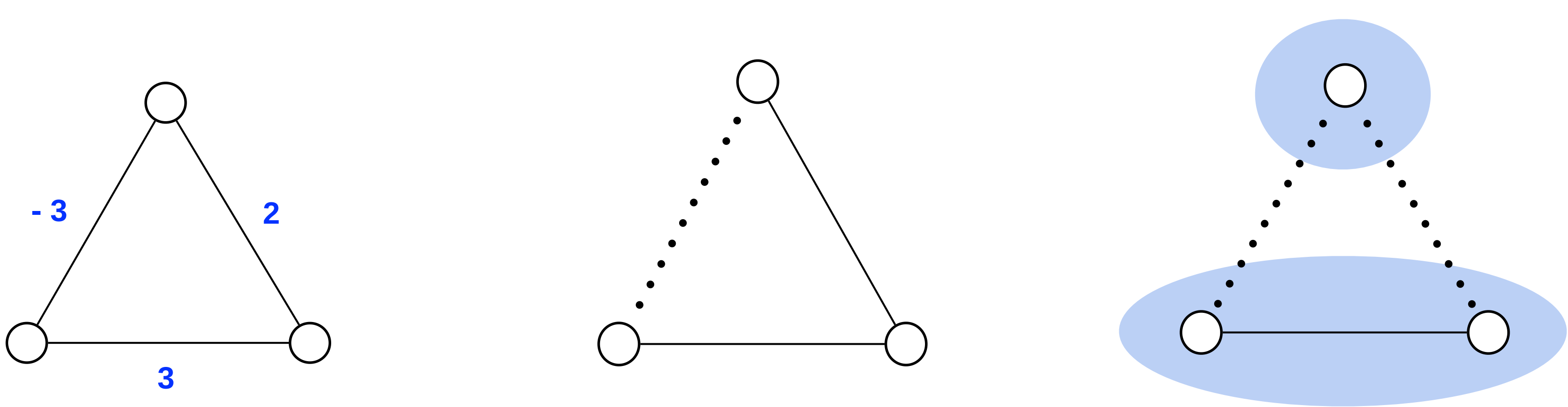


[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

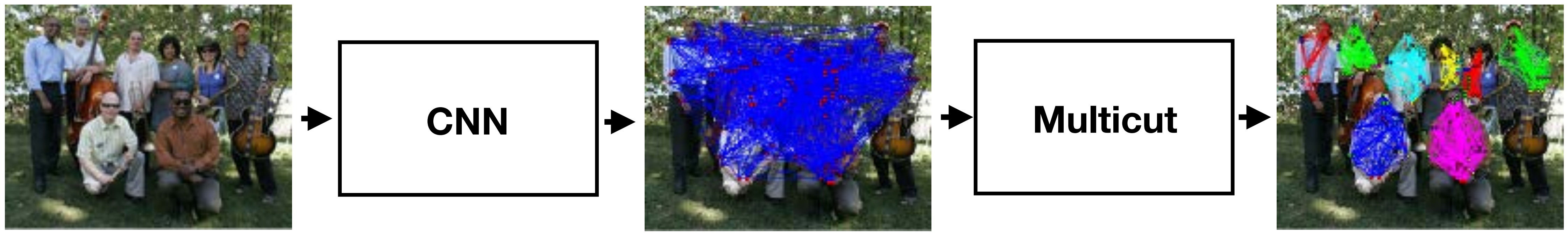
- Minimum Cost Multicut Problem [Chopra et al. The partition problem. *Mathematical Programming 1993*]

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \mathrm{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad .$$

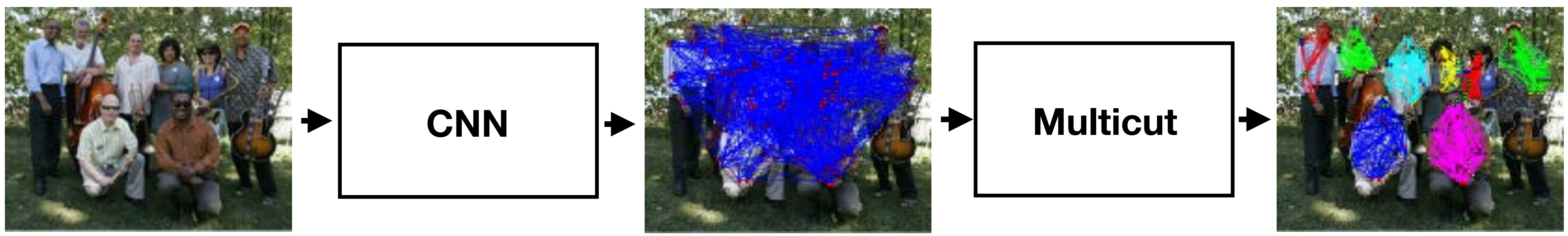


Multi-person Pose Estimation by Graph Decomposition



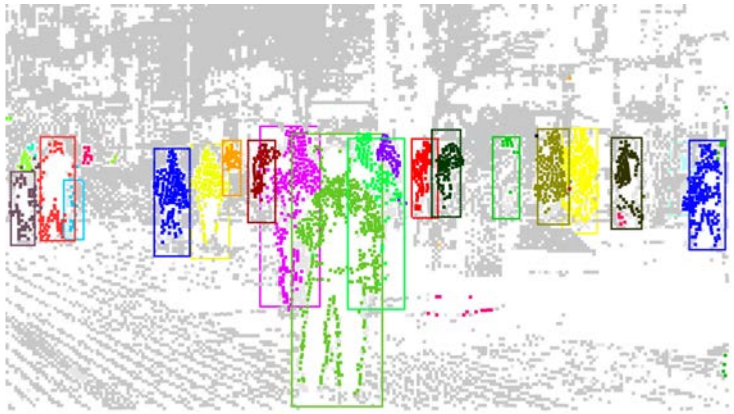
[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

Multi-person Pose Estimation by Graph Decomposition



[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

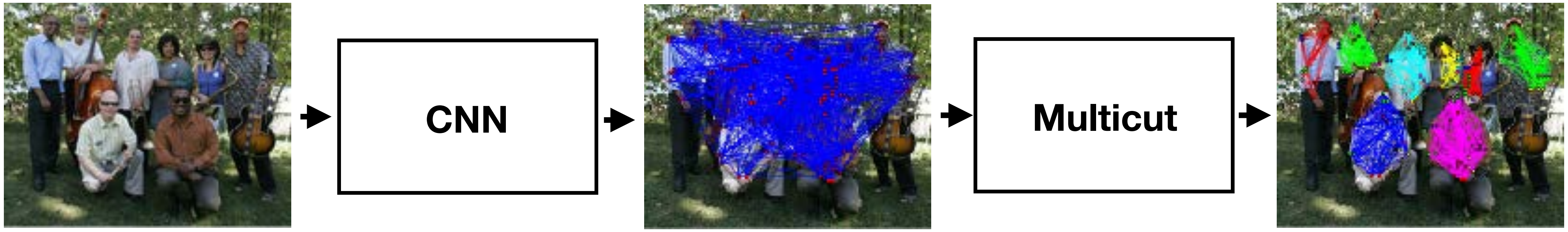
[Tang et al. CVPR 15, CVPR 2017]



[Keuper et al. TPAMI 2018]

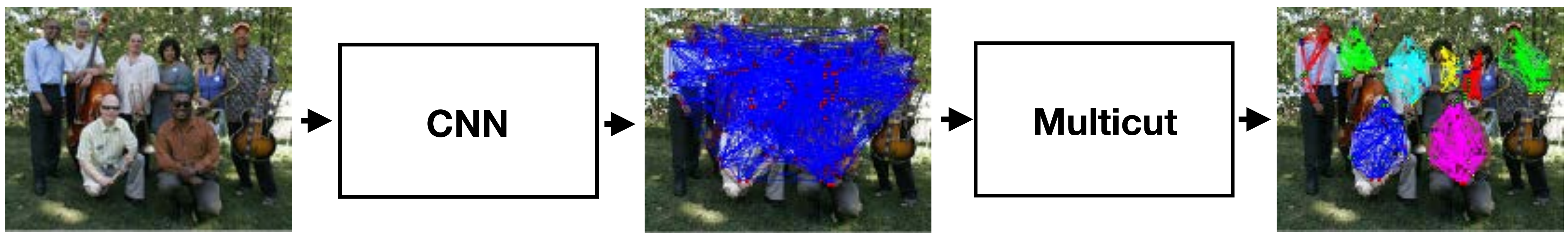
[Levinkov et al. CVPR2017]

Multi-person Pose Estimation by Graph Decomposition



[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

Multi-person Pose Estimation by Graph Decomposition

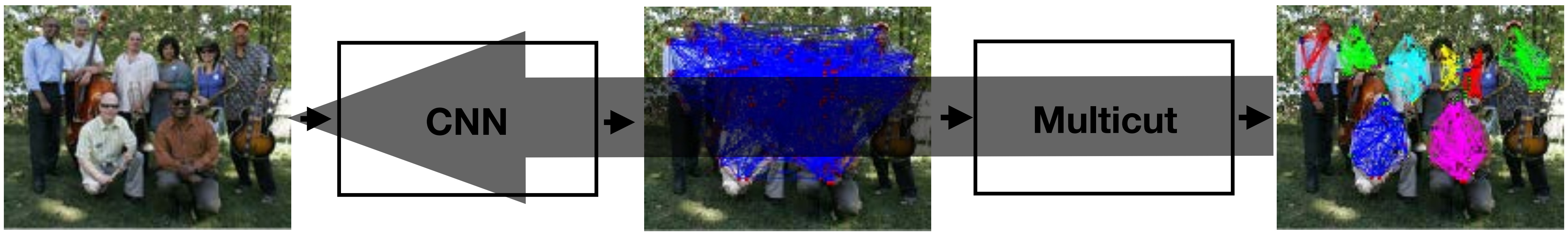


[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

- How to ***jointly*** learn the model parameters of Multicut and the weights of the front end CNNs?

- How to use the ***cycle consistency constraints*** as supervisory signals?

Multi-person Pose Estimation by Graph Decomposition



[Pishchulin et al CVPR 2016, Insafutdinov et al CVPR 2017, Levinkov et al CVPR 2017]

- How to ***jointly*** learn the model parameters of Multicut and the weights of the front end CNNs?

- How to use the ***cycle consistency constraints*** as supervisory signals?

End-to-end learning for multicut

- The minimum cost multicut problem

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \mathrm{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad .$$

End-to-end learning for multicut

- The minimum cost multicut problem

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e$$
$$\text{subject to} \quad \forall C \in \mathrm{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \ .$$

- An unconstrained binary multilinear problem

$$\min_{x \in \{0,1\}^E} \quad \sum_{e \in E} c_e \, x_e + K \sum_{C \in \mathrm{cc}(G)} \sum_{e \in C} x_e \prod_{e' \in C \setminus \{e\}} (1 - x_{e'}) \ .$$

End-to-end learning for multicut

- The minimum cost multicut problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \text{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \ .$$

- An unconstrained binary multilinear problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{C \in \text{cc}(G)} \sum_{e \in C} x_e \prod_{e' \in C \setminus \{e\}} (1 - x_{e'}) \ .$$

End-to-end learning for multicut

- The minimum cost multicut problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e$$

$$\text{subject to} \quad \forall C \in \mathrm{cc}(G) \, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad .$$

- An unconstrained binary multilinear problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{C \in \mathrm{cc}(G)} \sum_{e \in C} x_e \prod_{e' \in C \setminus \{e\}} (1 - x_{e'}) \quad .$$

End-to-end learning for multicut

- The minimum cost multicut problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e\, x_e$$

$$\text{subject to} \quad \forall C \in \mathrm{cc}(G)\, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \quad .$$

- An unconstrained binary multilinear problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e\, x_e + K \sum_{C \in \mathrm{cc}(G)} \sum_{e \in C} x_e \prod_{e' \in C \setminus \{e\}} (1 - x_{e'}) \quad .$$

- A binary cubic problem for the complete graph

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e\, x_e + K \sum_{\{u,v,w\} \in \binom{V}{3}} (x_{uv} \bar{x}_{vw} \bar{x}_{uw} + \bar{x}_{uv} x_{vw} \bar{x}_{uw} + \bar{x}_{uv} \bar{x}_{vw} x_{uw}) \quad .$$

End-to-end learning for multicut

- The minimum cost multicut problem

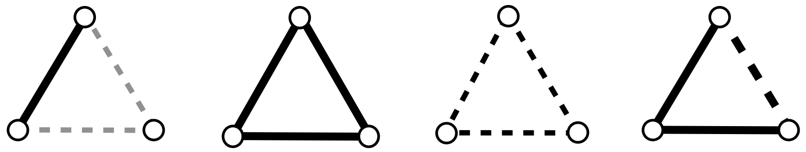
$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e$$
$$\text{subject to} \quad \forall C \in \mathrm{cc}(G)\, \forall e \in C : \quad x_e \leq \sum_{e' \in C \setminus \{e\}} x_{e'} \ .$$

- An unconstrained binary multilinear problem

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{C \in \mathrm{cc}(G)} \sum_{e \in C} x_e \prod_{e' \in C \setminus \{e\}} (1 - x_{e'}) \ .$$

- A binary cubic problem for the complete graph

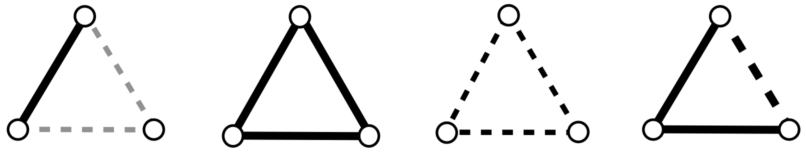
$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{\{u,v,w\} \in \binom{V}{3}} (x_{uv} \bar{x}_{vw} \bar{x}_{uw} + \bar{x}_{uv} x_{vw} \bar{x}_{uw} + \bar{x}_{uv} \bar{x}_{vw} x_{uw}) \ .$$



End-to-end learning for multicut

- A binary cubic problem for the complete graph

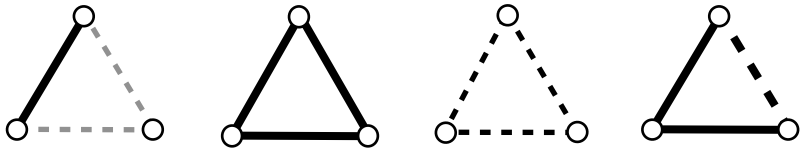
$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{\{u,v,w\} \in \binom{V}{3}} \left(x_{uv}\bar{x}_{vw}\bar{x}_{uw} + \bar{x}_{uv}x_{vw}\bar{x}_{uw} + \bar{x}_{uv}\bar{x}_{vw}x_{uw} \right) \; .$$



End-to-end learning for multicut

- A binary cubic problem for the complete graph

$$\min_{x\in\{0,1\}^E} \sum_{e\in E} c_e\, x_e + K \sum_{\{u,v,w\}\in\binom{V}{3}} (x_{uv}\bar{x}_{vw}\bar{x}_{uw} + \bar{x}_{uv}x_{vw}\bar{x}_{uw} + \bar{x}_{uv}\bar{x}_{vw}x_{uw}) \ .$$



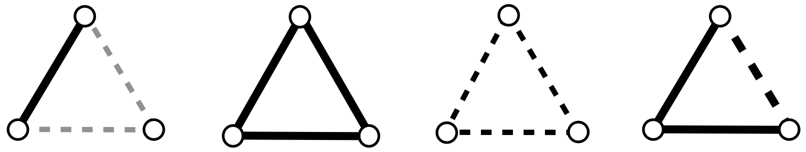
- Conditional Random Field

$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$

End-to-end learning for multicut

- A binary cubic problem for the complete graph

$$\min_{x\in\{0,1\}^E} \sum_{e\in E} c_e\, x_e + K \sum_{\{u,v,w\}\in\binom{V}{3}} \left(x_{uv}\bar{x}_{vw}\bar{x}_{uw} + \bar{x}_{uv}x_{vw}\bar{x}_{uw} + \bar{x}_{uv}\bar{x}_{vw}x_{uw}\right) .$$



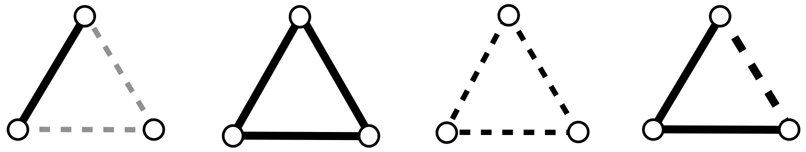
- Conditional Random Field

$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$

End-to-end learning for multicut

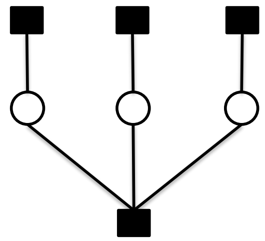
- A binary cubic problem for the complete graph

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{\{u,v,w\} \in \binom{V}{3}} \left(x_{uv}\bar{x}_{vw}\bar{x}_{uw} + \bar{x}_{uv}x_{vw}\bar{x}_{uw} + \bar{x}_{uv}\bar{x}_{vw}x_{uw} \right) .$$



- Conditional Random Field

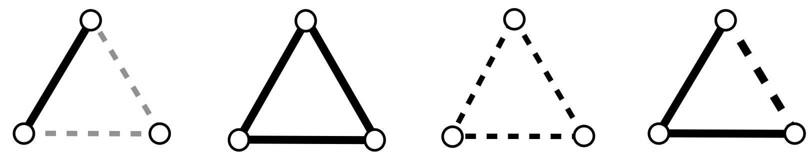
$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$



End-to-end learning for multicut

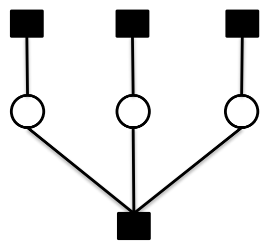
- A binary cubic problem for the complete graph

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{\{u,v,w\} \in \binom{V}{3}} \left(x_{uv} \bar{x}_{vw} \bar{x}_{uw} + \bar{x}_{uv} x_{vw} \bar{x}_{uw} + \bar{x}_{uv} \bar{x}_{vw} x_{uw} \right) .$$



- Conditional Random Field

$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$



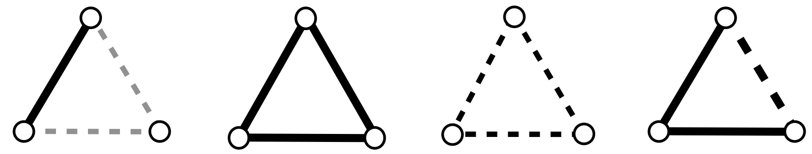
  - Pattern-based potential [Vineet et al. @ECCV 2012]

$$\psi_c^{Cycle}(\mathrm{x}_c) = \begin{cases} \gamma_{\mathbf{x}_c} & \text{if } \mathbf{x}_c \in \mathcal{P}_c \\ \gamma_{\max} & \text{otherwise} \end{cases}$$

End-to-end learning for multicut

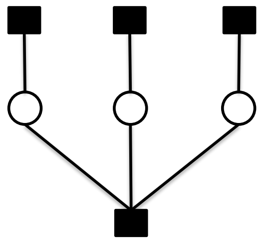
- A binary cubic problem for the complete graph

$$\min_{x \in \{0,1\}^E} \sum_{e \in E} c_e \, x_e + K \sum_{\{u,v,w\} \in \binom{V}{3}} \left(x_{uv} \bar{x}_{vw} \bar{x}_{uw} + \bar{x}_{uv} x_{vw} \bar{x}_{uw} + \bar{x}_{uv} \bar{x}_{vw} x_{uw} \right) .$$



- Conditional Random Field

$$E(x) = \sum_i \psi_i^U (x_i) + \sum_c \psi_c^{Cycle} (\mathrm{x}_c)$$

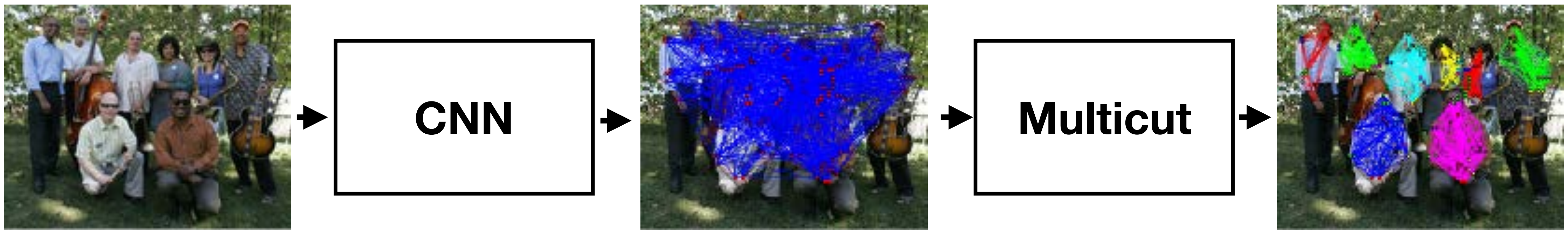


  - Pattern-based potential [Vineet et al. @ECCV 2012]

$$\psi_c^{Cycle}(\mathrm{x}_c) = \begin{cases} \gamma_{\mathbf{x}_c} & \text{if } \mathbf{x}_c \in \mathcal{P}_c \\ \gamma_{\max} & \text{otherwise} \end{cases}$$

  - Mean-field inference as RNN [Zheng et al. @ICCV 2015, Arnab et al. @ECCV 2016]

End-to-end learning for multicut



- Conditional Random Field

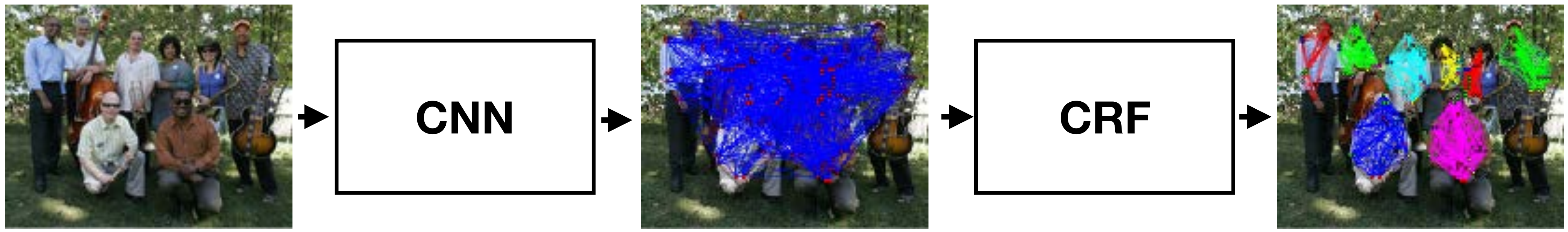
$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$

  - Pattern-based potential [Vineet et al. @ECCV 2012]

$$\psi_c^{Cycle}(\mathrm{x}_c) = \begin{cases} \gamma_{\mathbf{x}_c} & \text{if } \mathbf{x}_c \in \mathcal{P}_c \\ \gamma_{\max} & \text{otherwise} \end{cases}$$

  - Mean-field inference as RNN [Zheng et al. @ICCV 2015, Arnab et al. @ECCV 2016]

End-to-end learning for multicut



- Conditional Random Field

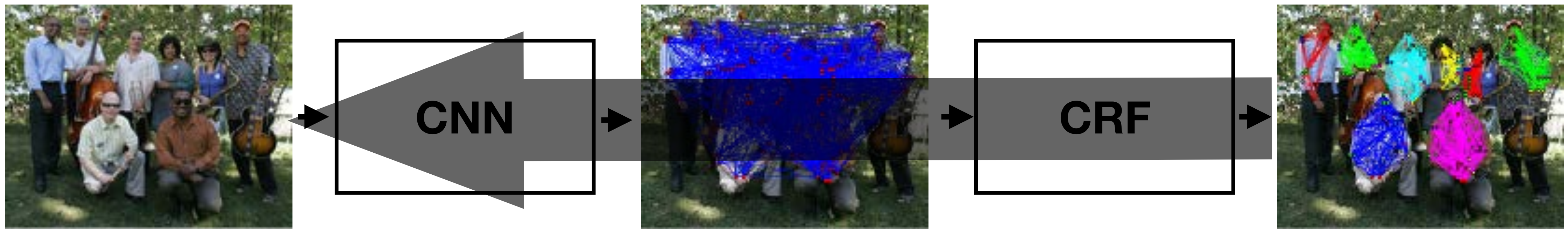
$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$

  - Pattern-based potential [Vineet et al. @ECCV 2012]

$$\psi_c^{Cycle}(\mathrm{x}_c) = \begin{cases} \gamma_{\mathbf{x}_c} & \text{if } \mathbf{x}_c \in \mathcal{P}_c \\ \gamma_{\max} & \text{otherwise} \end{cases}$$

  - Mean-field inference as RNN [Zheng et al. @ICCV 2015, Arnab et al. @ECCV 2016]

End-to-end learning for multicut



- Conditional Random Field

$$E(x) = \sum_i \psi_i^U(x_i) + \sum_c \psi_c^{Cycle}(\mathrm{x}_c)$$

  - Pattern-based potential [Vineet et al. @ECCV 2012]

$$\psi_c^{Cycle}(\mathrm{x}_c) = \begin{cases} \gamma_{\mathbf{x}_c} & \text{if } \mathbf{x}_c \in \mathcal{P}_c \\ \gamma_{\max} & \text{otherwise} \end{cases}$$

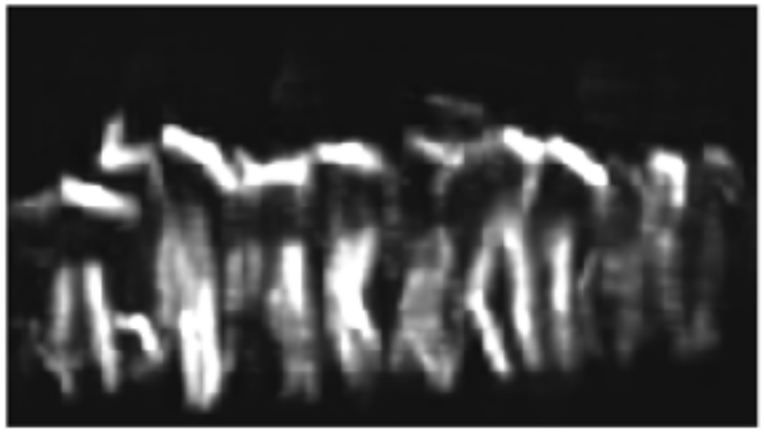
  - Mean-field inference as RNN [Zheng et al. @ICCV 2015, Arnab et al. @ECCV 2016]

Experiments

Experiments

Input

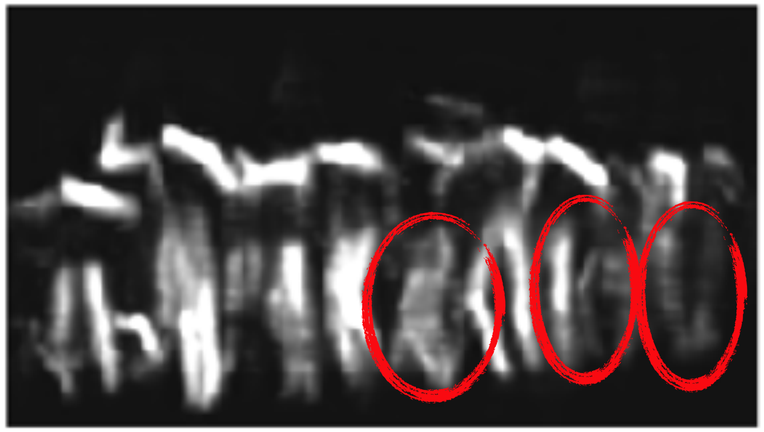
Part affinity field from OpenPose



Experiments

Input

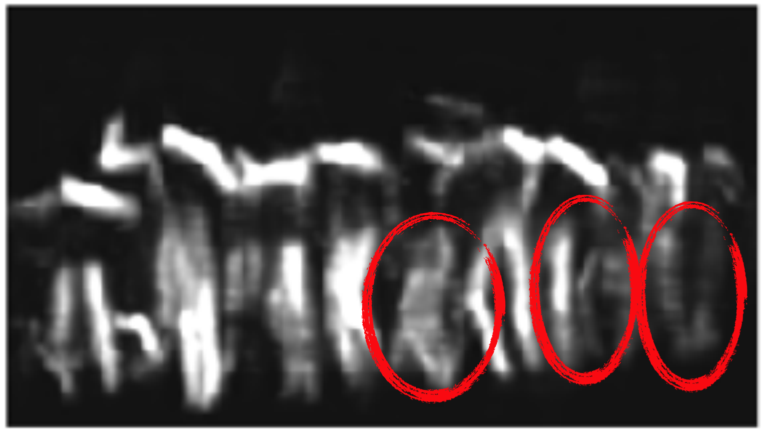
Part affinity field from OpenPose



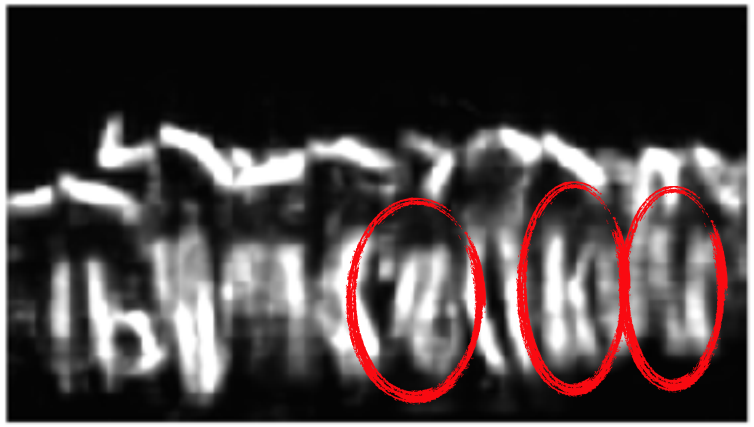
Experiments

Input

Part affinity field from OpenPose



with end-to-end training



Experiments

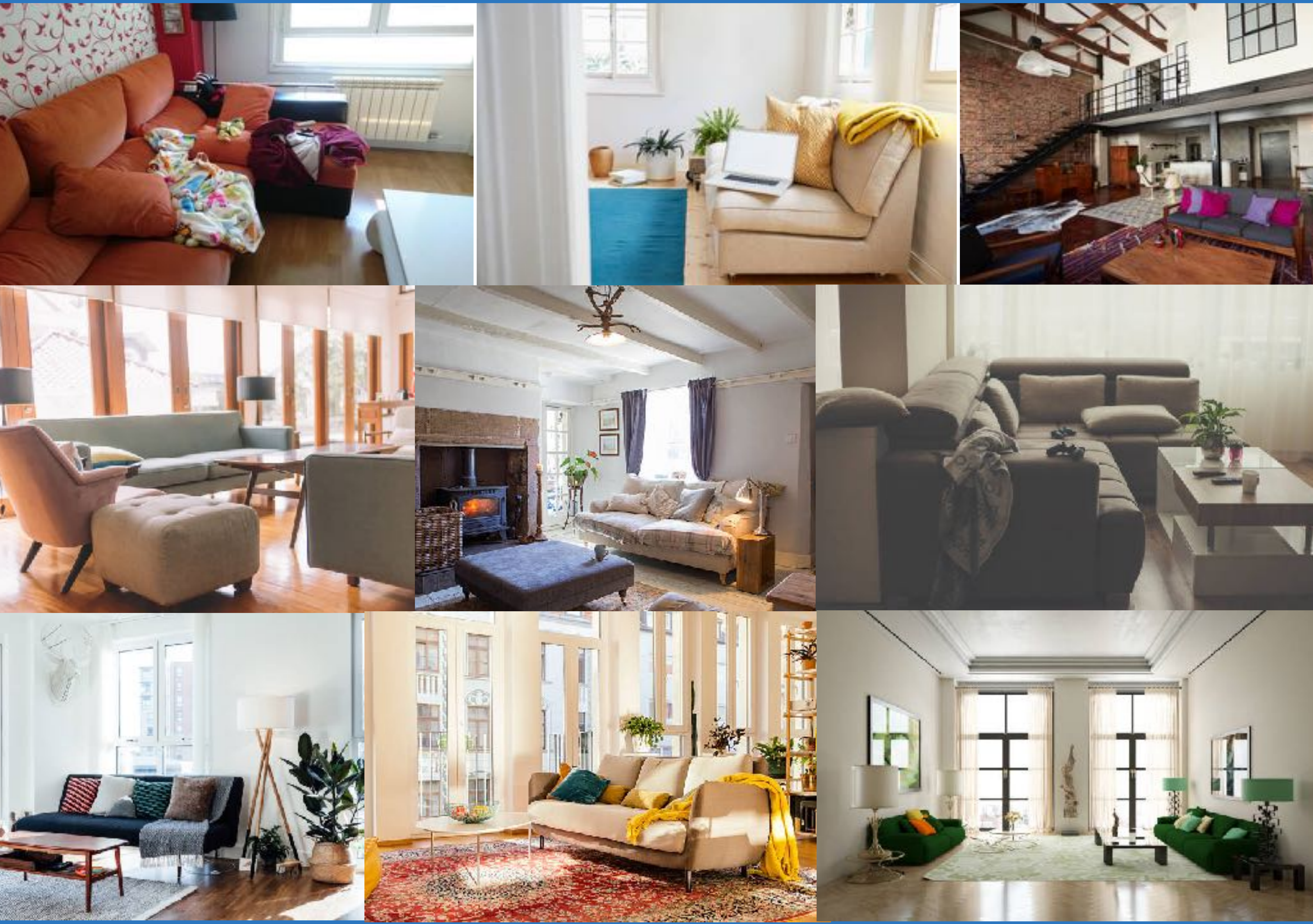
Experiments

*End-to-end Learning for Graph Decomposition. Song, Andres, Black, Hilliges, **Tang.** ICCV 2019*

Part 2: Learning to generate humans

Part 2: Learning to generate humans

Seeing People in Images without People



Probabilistic 3D Body Generation in Images

Probabilistic 3D Body Generation in Images

- Scene and human body representation

Probabilistic 3D Body Generation in Images

- Scene and human body representation

Probabilistic 3D Body Generation in Images

- Scene and human body representation
    - Estimate 3D body

Probabilistic 3D Body Generation in Images

- Scene and human body representation

- Estimate 3D body

- Estimate semantic segmentation

Probabilistic 3D Body Generation in Images

- Scene and human body representation

- Estimate 3D body

- Estimate semantic segmentation

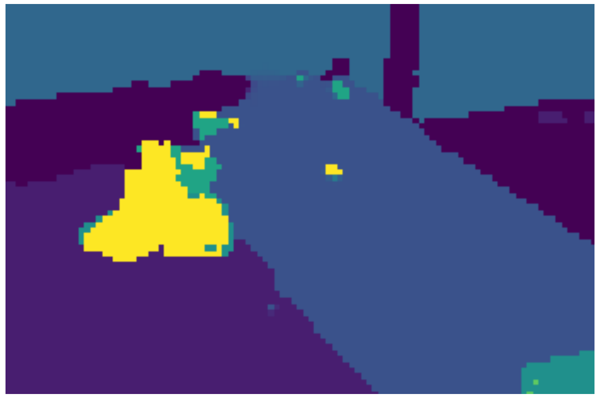
- Estimate depth

Probabilistic 3D Body Generation in Images

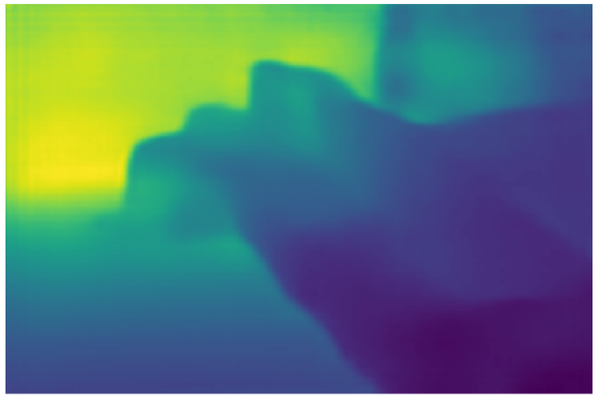
- Scene and human body representation

- Estimate 3D body

- Estimate semantic segmentation



- Estimate depth



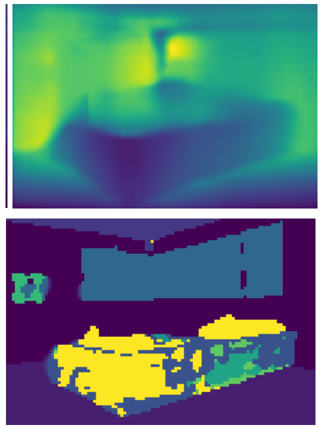
Probabilistic 3D Body Generation in Images

- Conditional Human body generation
    - Conditional module: Environment net

Probabilistic 3D Body Generation in Images

- Conditional Human body generation

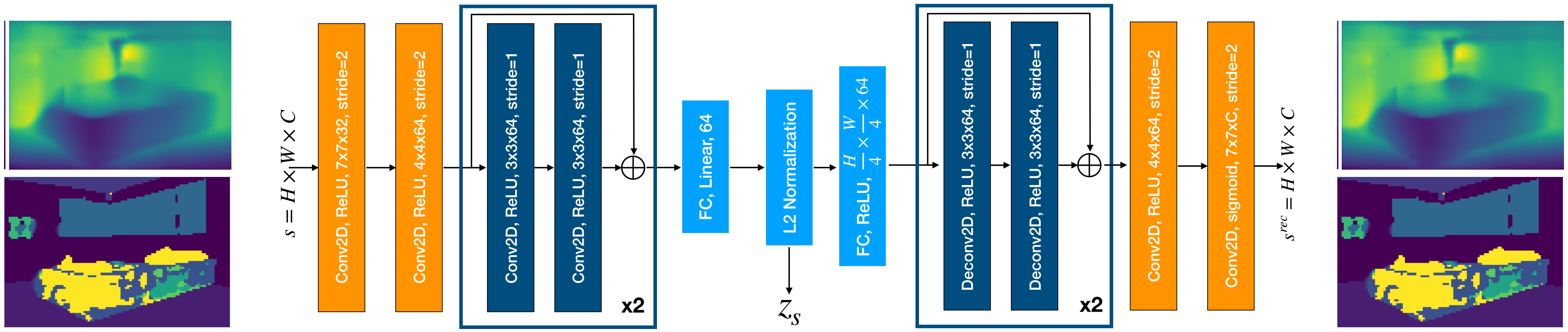
  - Conditional module: Environment net



Probabilistic 3D Body Generation in Images

- Conditional Human body generation

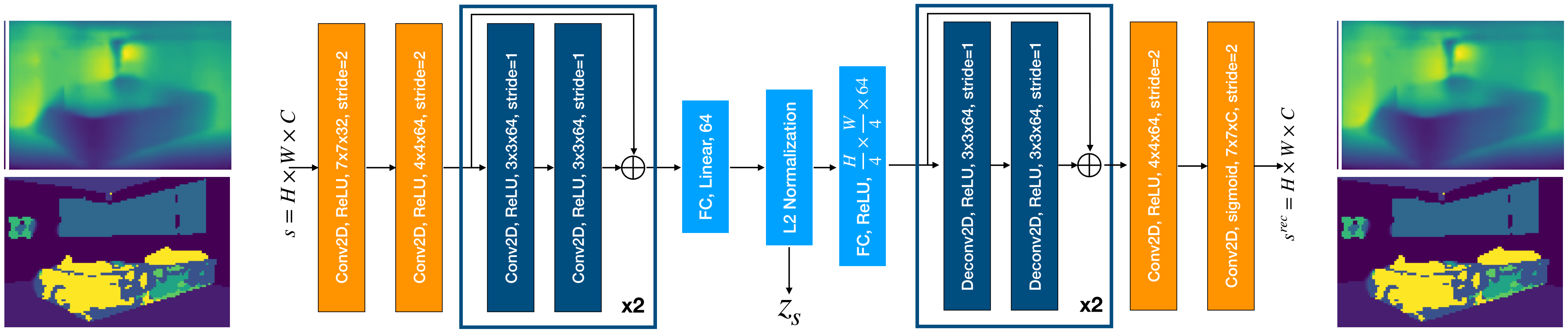
  - Conditional module: Environment net



Probabilistic 3D Body Generation in Images

- Conditional Human body generation

  - Conditional module: Environment net

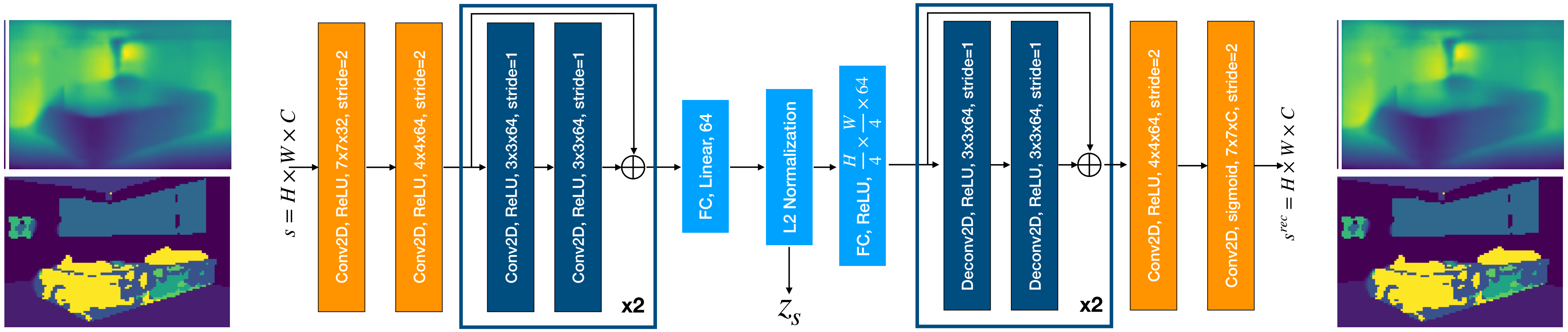


  - Conditional variational autoencoder

Probabilistic 3D Body Generation in Images

- Conditional Human body generation

  - Conditional module: Environment net

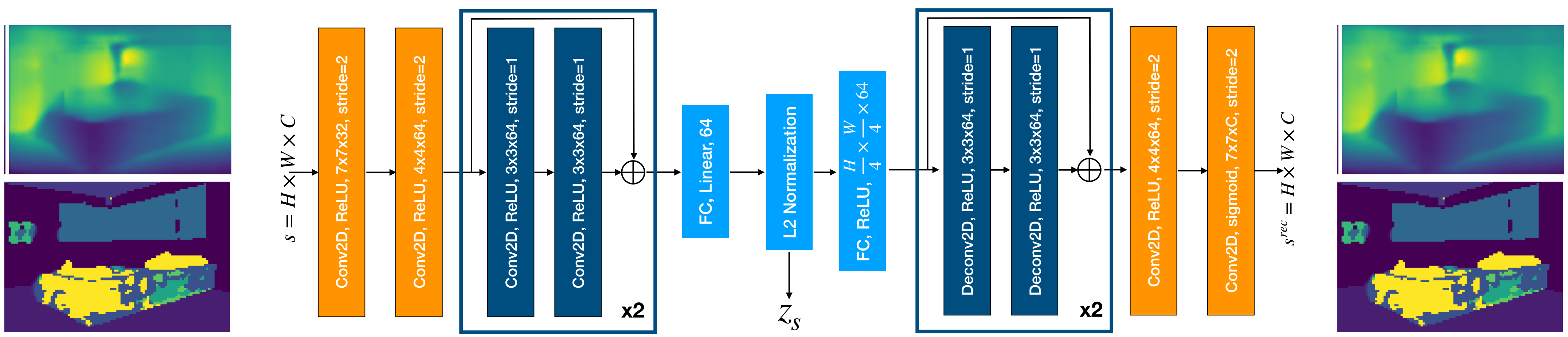


  - Conditional variational autoencoder

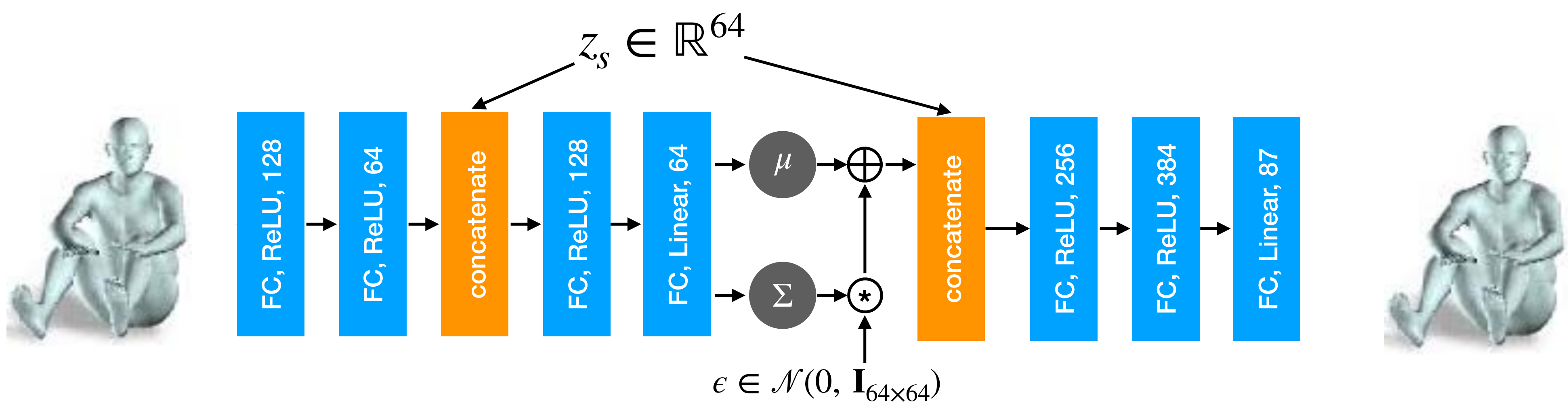
Probabilistic 3D Body Generation in Images

- Conditional Human body generation

  - Conditional module: Environment net

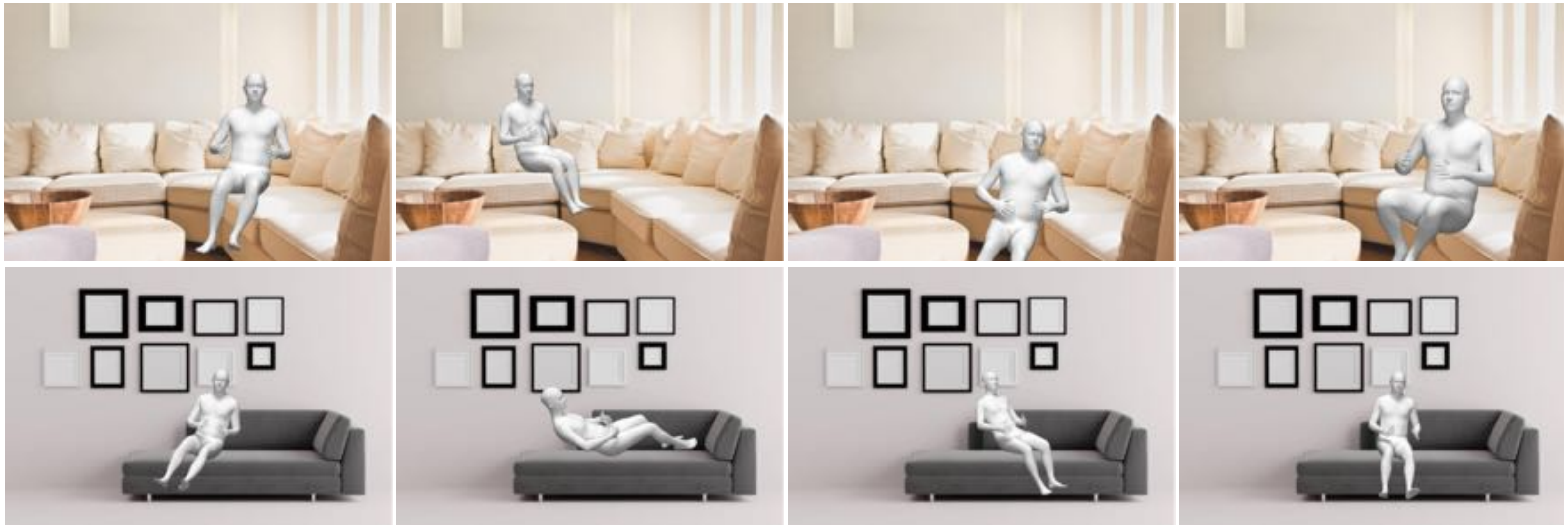


  - Conditional variational autoencoder

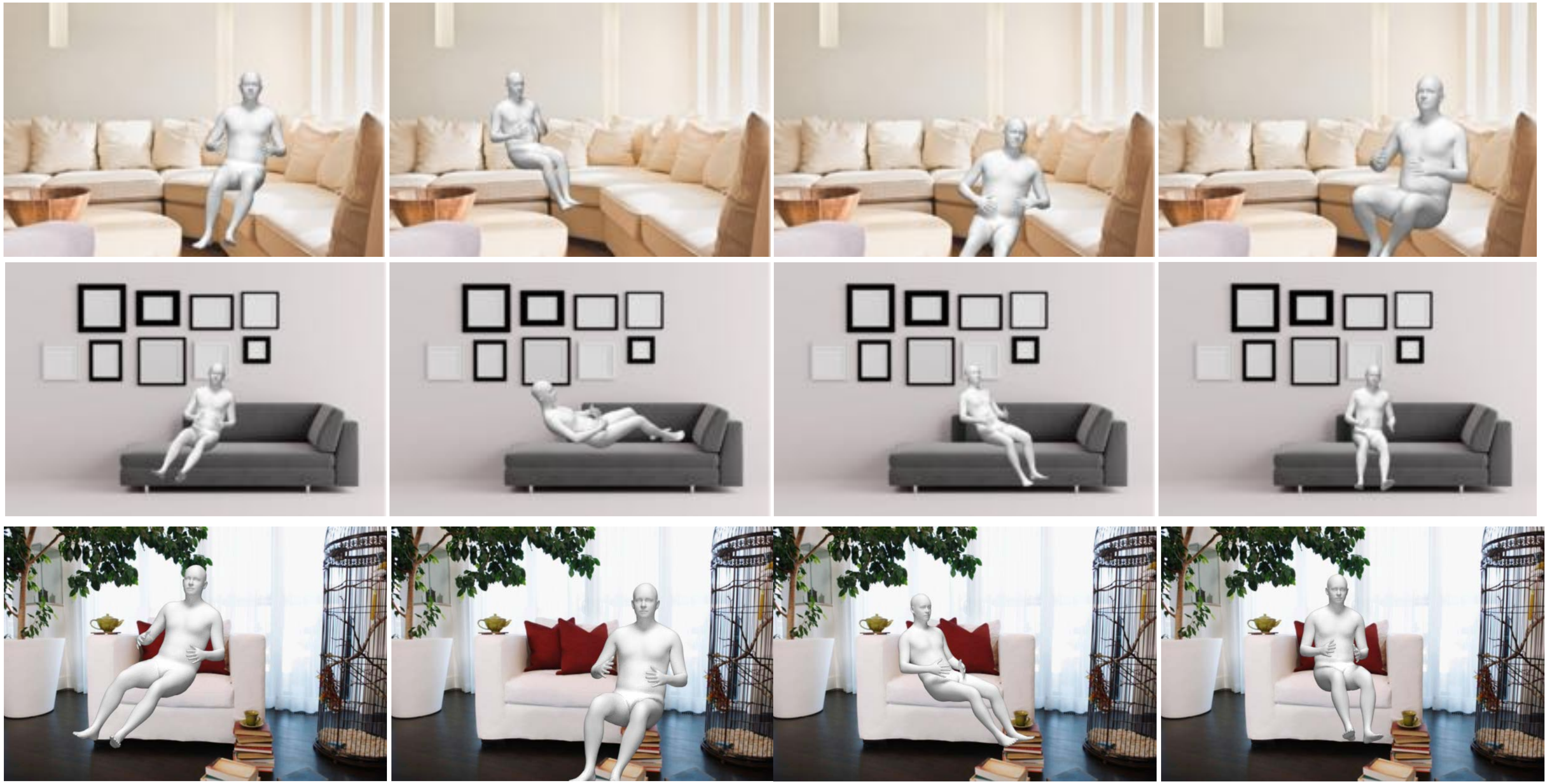


Result

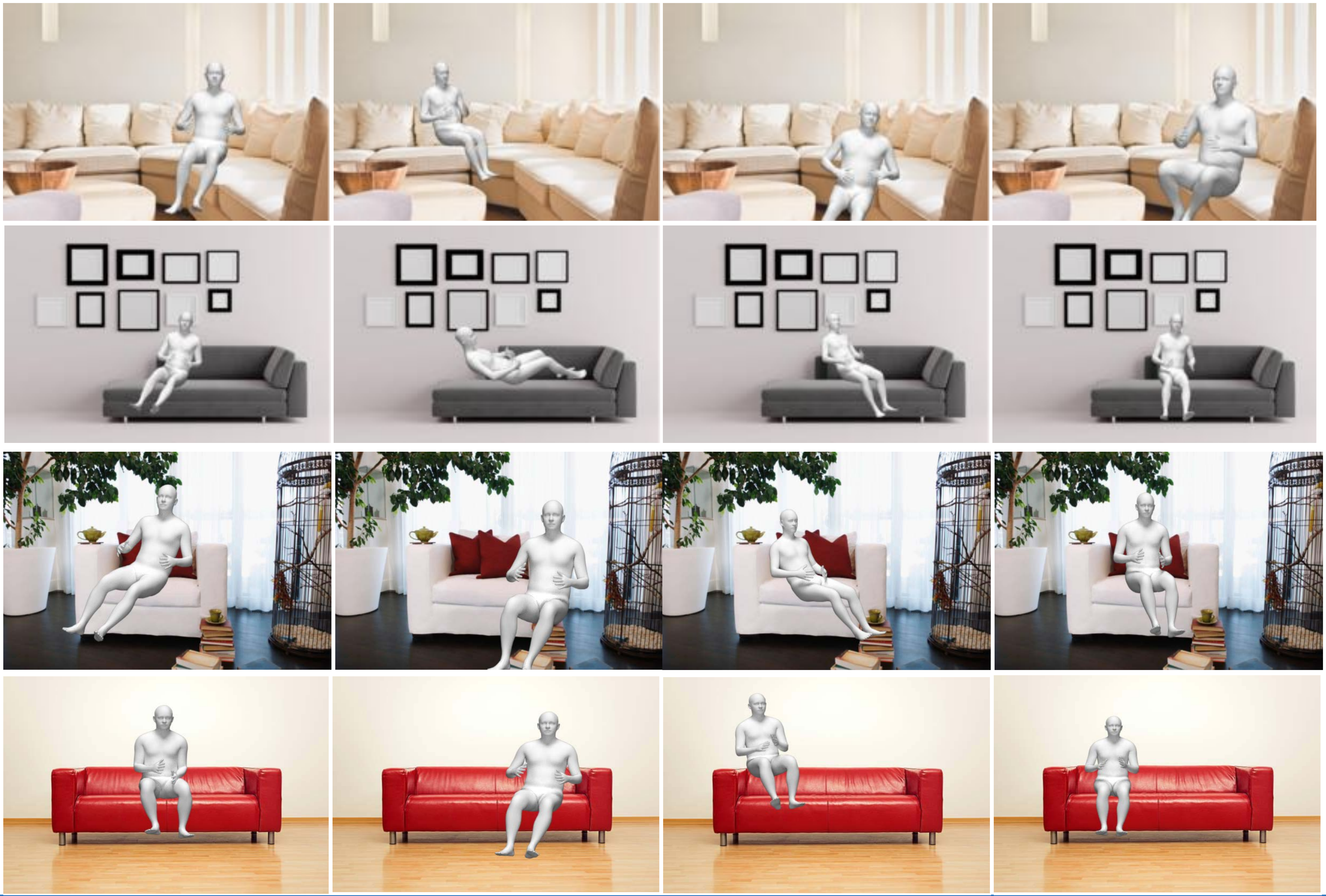
Result



Result



Result



Result

Result

- MPI-INF-3DHP sitting sequence

Result

- MPI-INF-3DHP sitting sequence

Threshold (mm)	20	40	60	80	100
VPoser [Pavlakos et al. @CVPR 2019]	13.15	34.92	54.02	64.34	71.60
Ours	**18.14**	**43.82**	**65.08**	**76.87**	**82.48**

Result

- MPI-INF-3DHP sitting sequence

Threshold (mm)	20	40	60	80	100
VPoser [Pavlakos et al. @CVPR 2019]	13.15	34.92	54.02	64.34	71.60
Ours	**18.14**	**43.82**	**65.08**	**76.87**	**82.48**

- Occlusion handeling


Result

- MPI-INF-3DHP sitting sequence

Threshold (mm)	20	40	60	80	100
VPoser [Pavlakos et al. @CVPR 2019]	13.15	34.92	54.02	64.34	71.60
Ours	**18.14**	**43.82**	**65.08**	**76.87**	**82.48**

- Occlusion handeling

- Failure cases

Result

- MPI-INF-3DHP sitting sequence

Threshold (mm)	20	40	60	80	100
VPoser [Pavlakos et al. @CVPR 2019]	13.15	34.92	54.02	64.34	71.60
Ours	**18.14**	**43.82**	**65.08**	**76.87**	**82.48**

- Occlusion handeling

- Failure cases

Result

- MPI-INF-3DHP sitting sequence

Threshold (mm)	20	40	60	80	100
VPoser [Pavlakos et al. @CVPR 2019]	13.15	34.92	54.02	64.34	71.60
Ours	**18.14**	**43.82**	**65.08**	**76.87**	**82.48**

- Occlusion handeling

- Failure cases

Thank you!