

# Alpparel: A Multimodal Foundation Model for Digital Garments

Kiyohiro Nakayama\*, Jan Ackmann\*, Timur Kesdogan\*, Yang Zheng, Maria Korosteleva, Leonidas Guibas, Olga Sorkine-Hornung, Guandao Yang, Gordon Wetzstein



ETH zürich



# Fashion is an important part of life



# Fashion Design is a Multistage Process



# Fashion Design is a Multistage Process

A fashion year consists of two seasons: **Spring** & **Summer**, **Autumn** & **Winter**.



# A Typical Design Cycle Consists of

**Ideation:** designers come up with a theme for the season.

# A Typical Design Cycle Consists of

**Ideation:** designers come up with a theme for the season.

Mood Board



# A Typical Design Cycle Consists of

**Ideation:** designers come up with a theme for the season.

**Sketching:** garments' designs are sketched on paper.

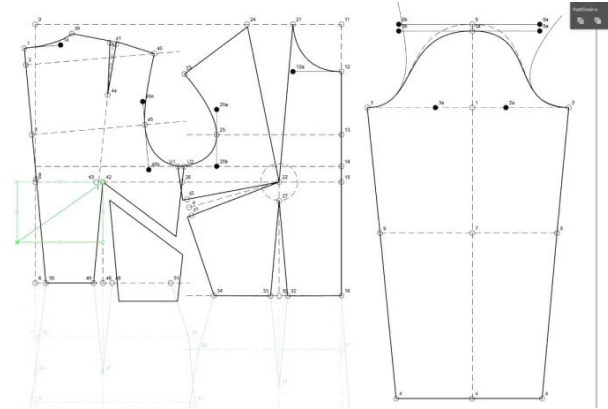


# A Typical Design Cycle Consists of

**Ideation:** designers come up with a theme for the season.

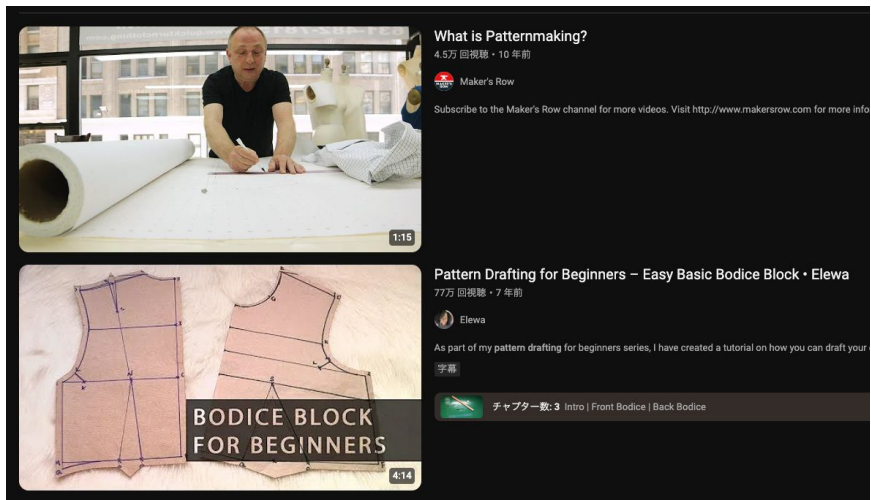
**Sketching:** garments' designs are sketched on paper.

**Pattern Making:** sketches are made into panels for sizing and production.



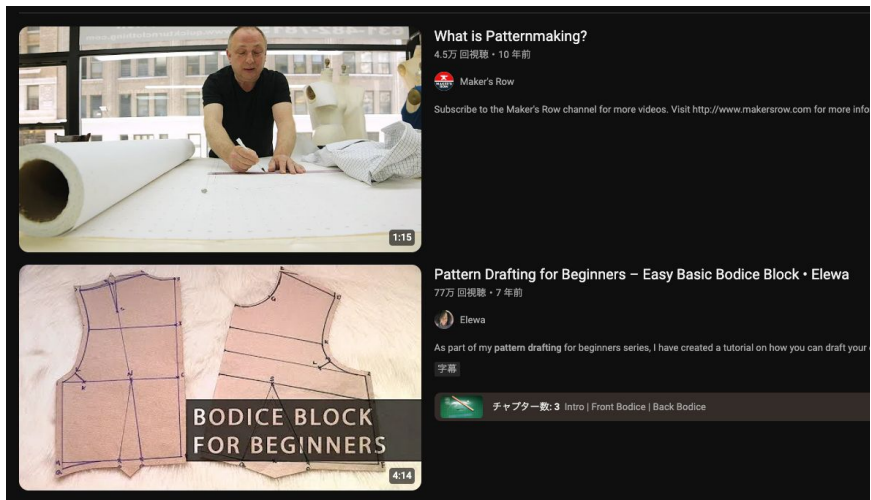
# Digitized Fashion Design

- Pattern making, in particular, is time-consuming, requires professional skills, and repetitive.



# Digitized Fashion Design

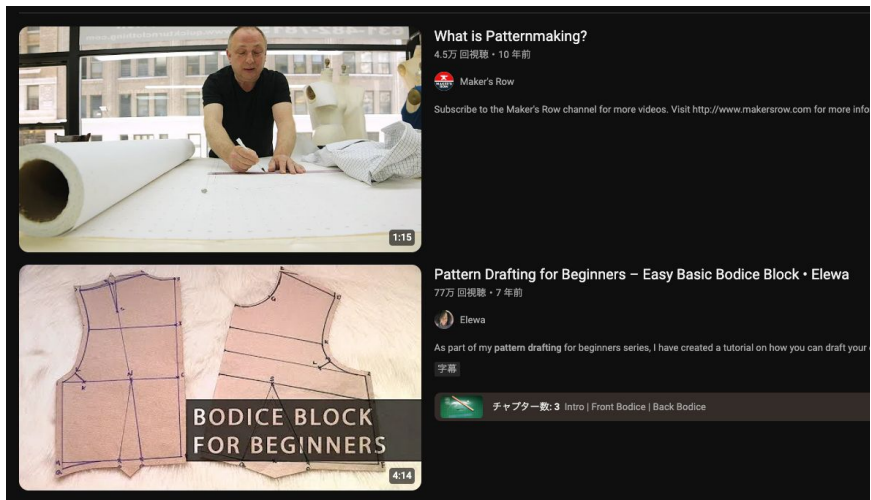
- Pattern making, in particular, is time-consuming, requires professional skills, and repetitive.
- Digital tools are developed to make this more convenient, but they are also difficult to master.



# Digitized Fashion Design

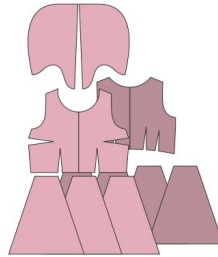
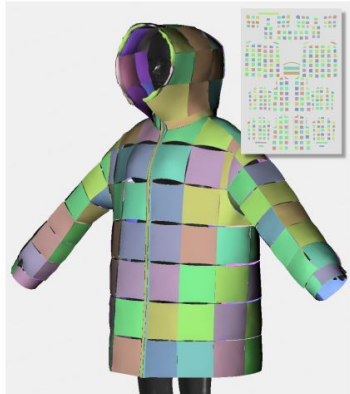
- Pattern making, in particular, is time-consuming, requires professional skills, and repetitive.
- Digital tools are developed to make this more convenient, but they are also difficult to master.

**Can we use generative models to automate this process?**



# Pattern Making 101: What is a Sewing Pattern

- A sewing pattern describes how fabrics should be stitched together to make a garment.
- A sewing pattern consists of **2D panels** and **how to stitch them together**.
- For digital draping, **3D rigid transformations** of 2D panels are also required.



Panel meshes



Box mesh



Simulated result

While sewing patterns are complex in practice...

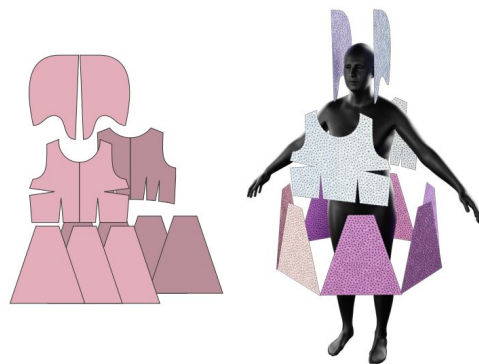
For us, sewing patterns are a set of 2D panels  $P_i$ ,  $i = 1, \dots, n$ .

# While sewing patterns are complex in practice...

For us, sewing patterns are a set of 2D panels  $P_i$ ,  $i = 1, \dots, n$ .

Each  $P_i$  consists of

1. (optionally) A name,
2. A 3D rigid transformation  $R_i$  &  $T_i$ ,
3. An ordered set of edges  $(e_i^1, e_i^2, \dots, e_i^{m_i})$ , that are either straight lines, cubic bezier curves, or arcs.
4. Vertices  $\{v_i^s\}_{s=1, \dots, m_i}$  and control points  $\{c_i^k\}_{k=1, \dots, m_i}$ .



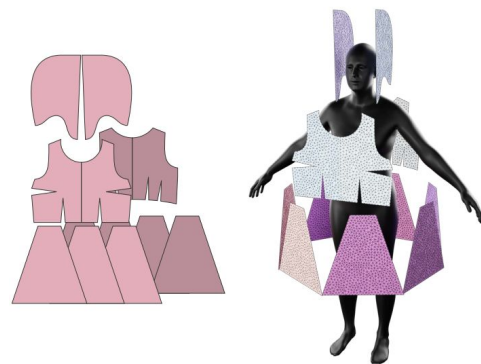
# While sewing patterns are complex in practice...

For us, sewing patterns are a set of 2D panels  $P_i$ ,  $i = 1, \dots, n$ .

Each  $P_i$  consists of

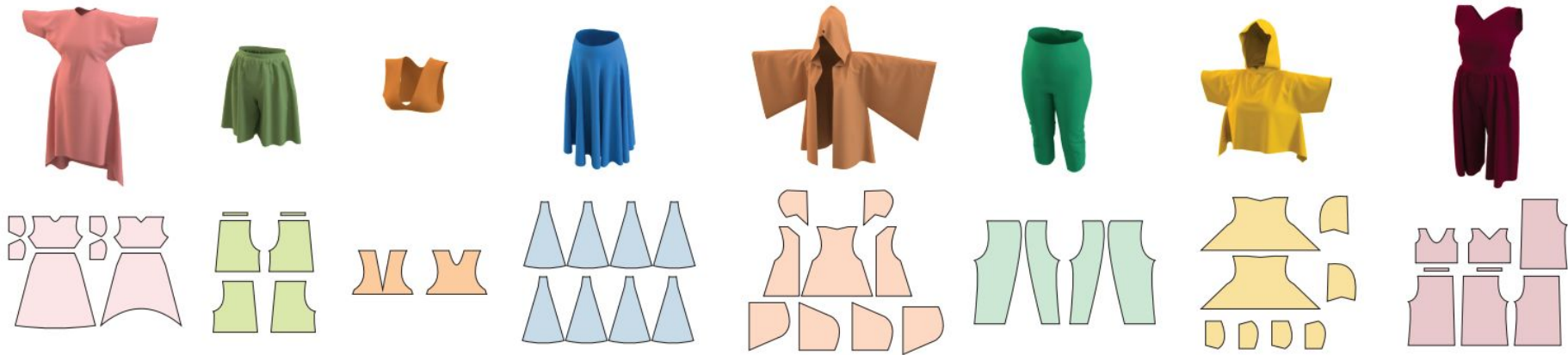
1. (optionally) A name,
2. A 3D rigid transformation  $R_i$  &  $T_i$ ,
3. An ordered set of edges  $(e_i^1, e_i^2, \dots, e_i^{m_i})$ , that are either straight lines, cubic bezier curves, or arcs.
4. Vertices  $\{v_i^s\}_{s=1, \dots, m_i}$  and control points  $\{c_i^k\}_{k=1, \dots, m_i}$ .

A set of edge pairs  $\{e_i^{k_i}, e_j^{l_j}\}$  indicates if two edges are stitched together (both within and across panels).



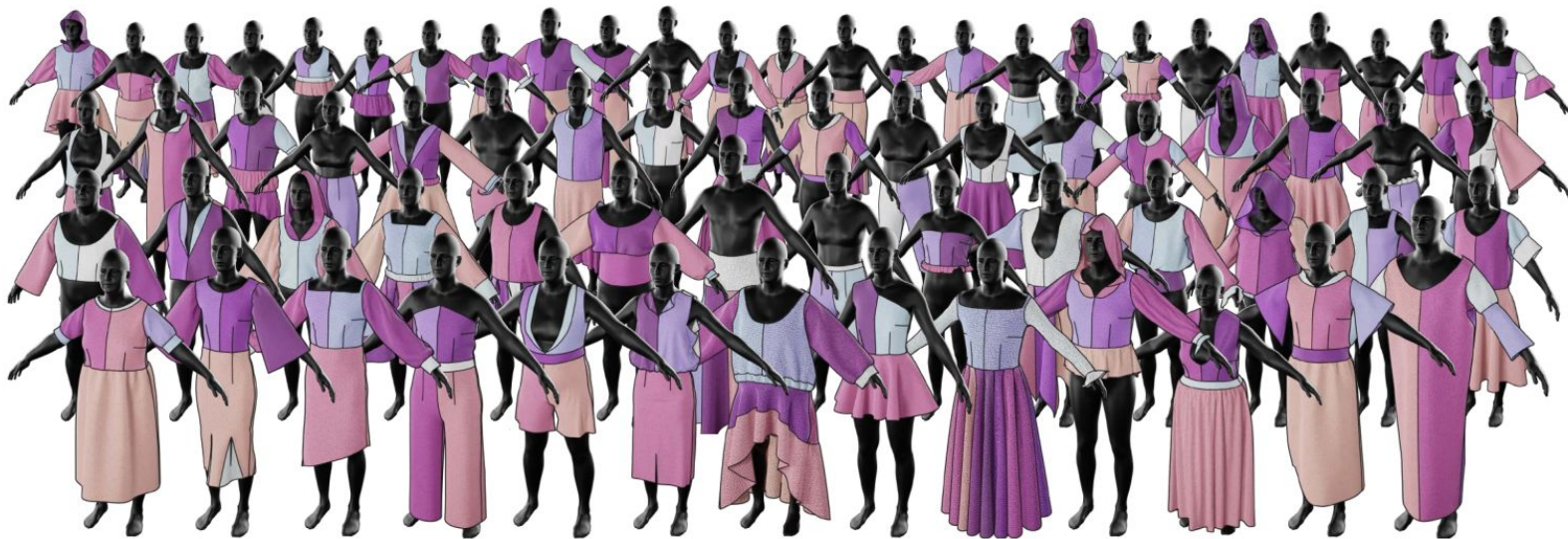
# Data

- [Maria & Lee, 2021] is relatively easy, with only 12 types of garments in total. This is what existing works use (~20k garments).

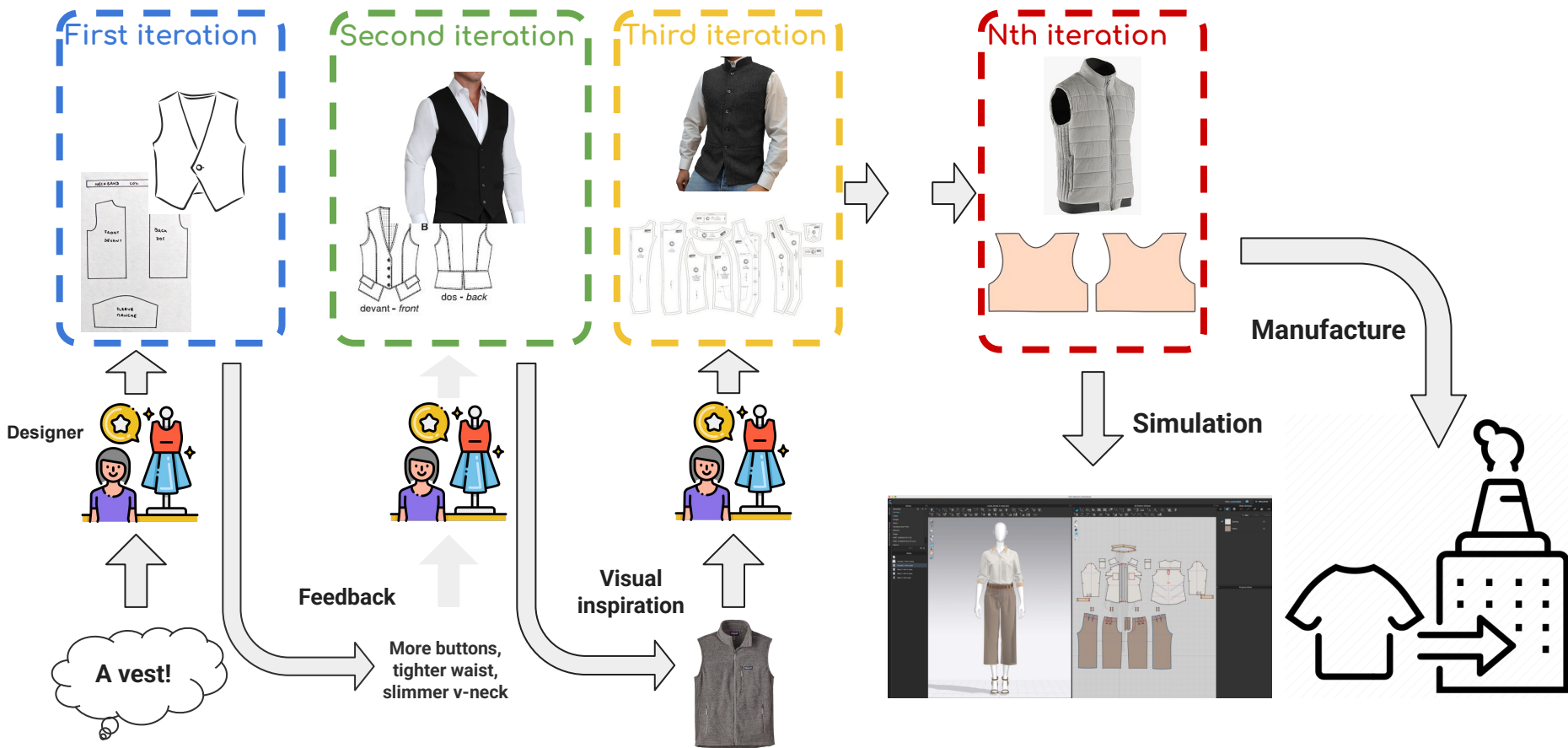


# Data

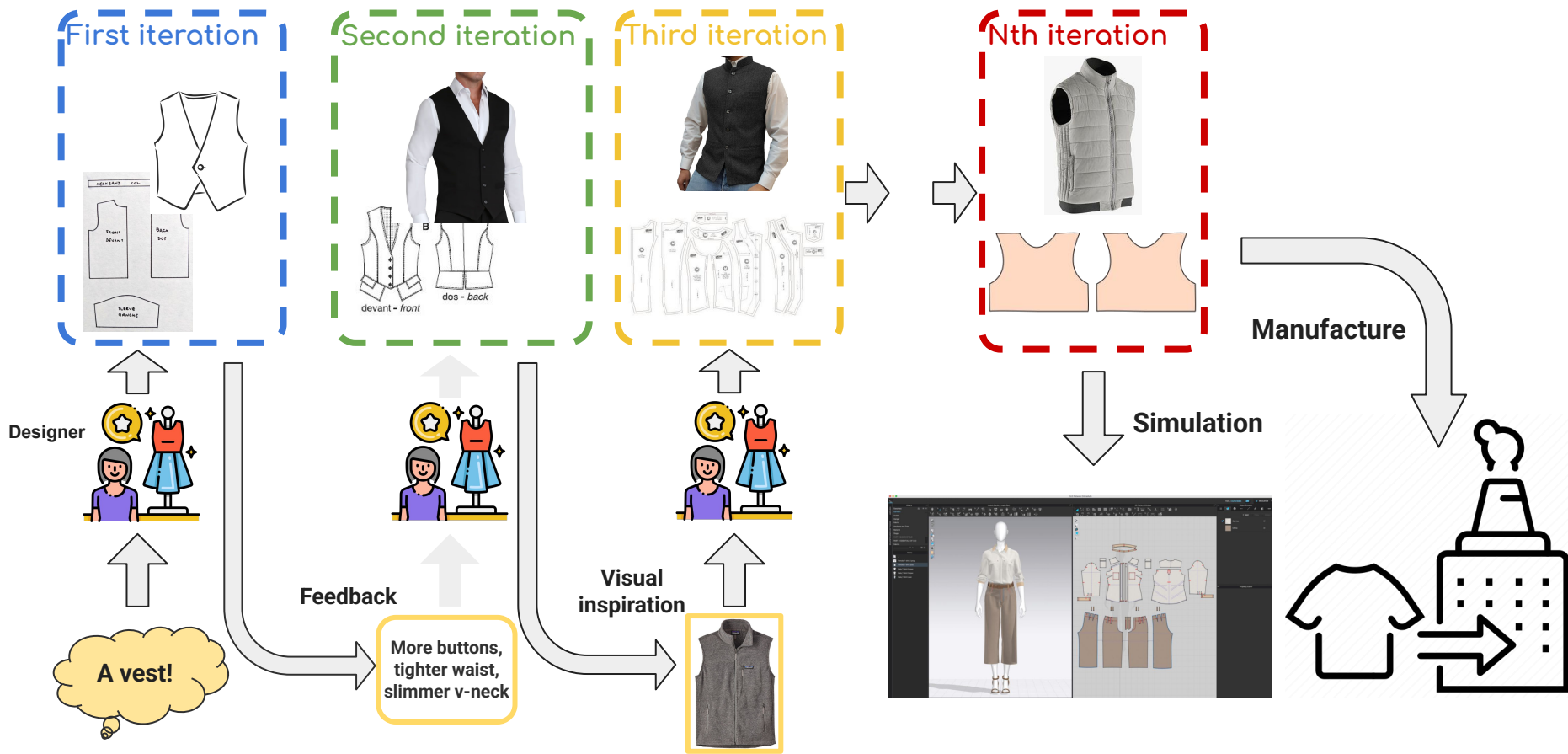
- [Maria & Lee, 2021] is relatively easy, with only 12 types of garments in total. This is what existing works use (~20k garments).
- Recently, a sewing pattern dataset, GarmentCodeData is published with more variety and complexity (~130k garments)



# Pattern Making is an Iterative Process



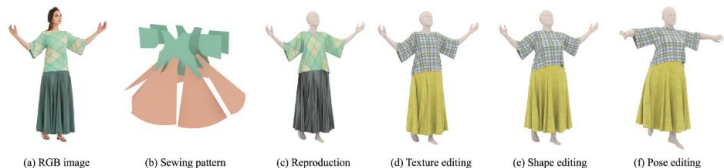
# Pattern Making is an Iterative Process



# Existing Works — Pattern Generation

Existing pattern generation methods are mode specific.

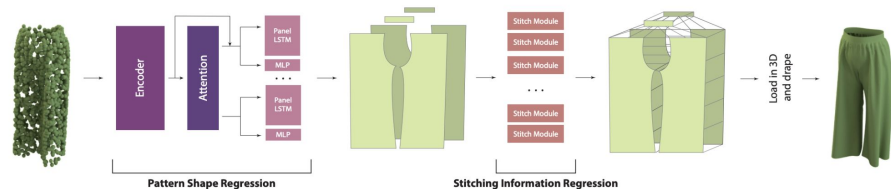
## Image → Pattern



Given a single RGB image of a clothed human (a), the proposed algorithm can accurately recover the underlying garment sewing pattern (b), leading to wide applications in virtual/augmented reality, for example, 3D garment mesh reconstruction (c) and 3D garment editing in terms of garment texture (d), human shape (e), and human pose (f).

[SewFormer, SigAisa 2023]

## Point Cloud → Pattern



[NeuralTailor, Siggraph 2022]

## Text → Pattern

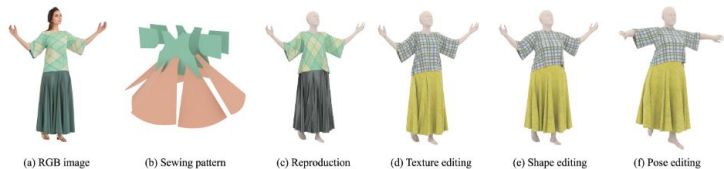


[DressCode, Siggraph 2024]

# Existing Works — Pattern Generation

Existing pattern generation methods are mode specific.

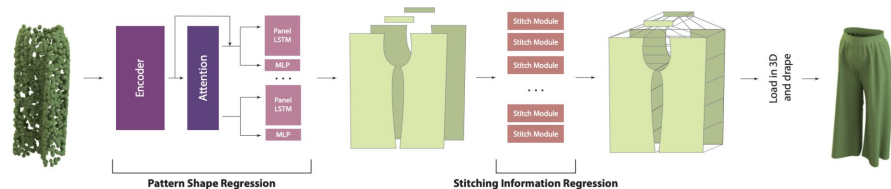
## Image → Pattern



Given a single RGB image of a clothed human (a), the proposed algorithm can accurately recover the underlying garment sewing pattern (b), leading to wide applications in virtual/augmented reality, for example, 3D garment mesh reconstruction (c) and 3D garment editing in terms of garment texture (d), human shape (e), and human pose (f).

[SewFormer, SigAisa 2023]

## Point Cloud → Pattern



[NeuralTailor, Siggraph 2022]

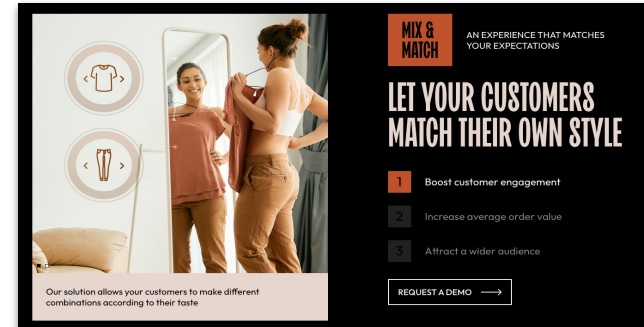
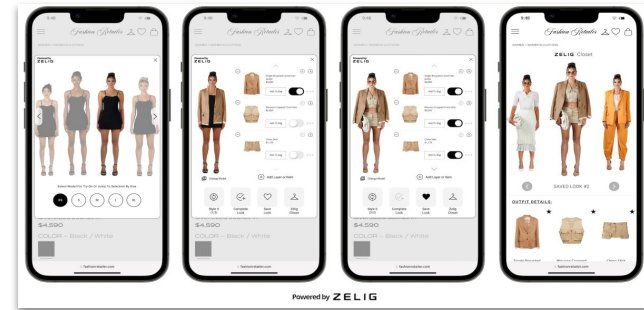
## Text → Pattern



[DressCode, Siggraph 2024]

Cannot generalize to other modalities.

# Related Works — Virtual Try-on Companies



- Virtual Try-on apps **do not** output manufacturable garments.
- Suitable for consumers, but hard to fit into existing design-manufacturing pipeline of clothing industry.

# Alpparel: Large Multimodal Generative Model for Digital Garments

## What we want:

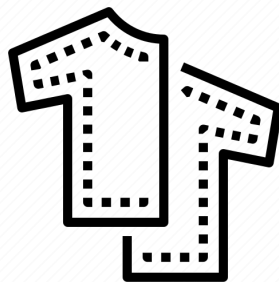
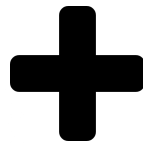
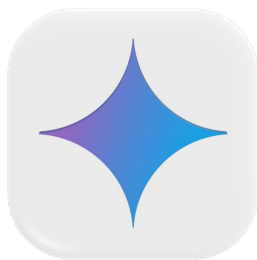
1. Generate sewing patterns from multimodal inputs.
2. Enable method's reasoning ability for iterative design.

# Alpparel: Large Multimodal Generative Model for Digital Garments

## What we want:

1. Generate sewing patterns from multimodal inputs.
2. Enable method's reasoning ability for iterative design.

Approach: we finetune a pre-trained Multimodal LLM (MLLM) on the task of multimodal sewing pattern generation.

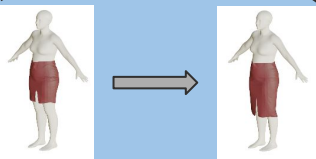


# Road Map

## Data:

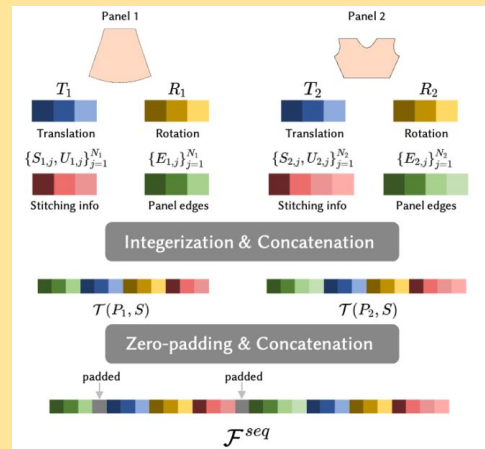
How do we build a large scale multimodal dataset of sewing patterns?

It's my best friend's wedding! What should I wear?



## Representation:

How do we represent sewing pattern for MLLMs?

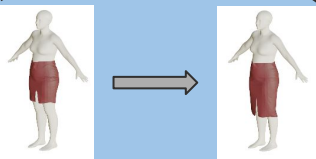


# Road Map

## Data:

How do we build a large scale multimodal dataset of sewing patterns?

It's my best friend's wedding! What should I wear?



## Representation:

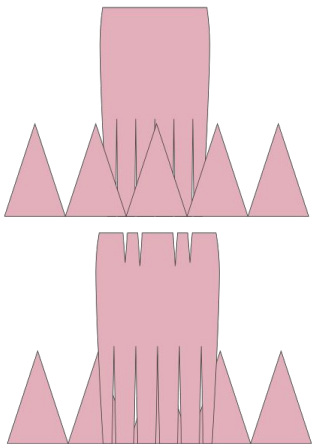
How do we represent sewing pattern for MLLMs?



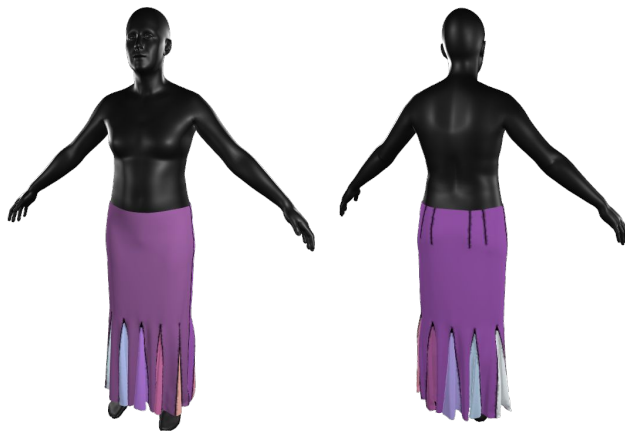
# Annotation Curation

- GarmentCodeData contains large-scale sewing patterns.

Sewing Pattern



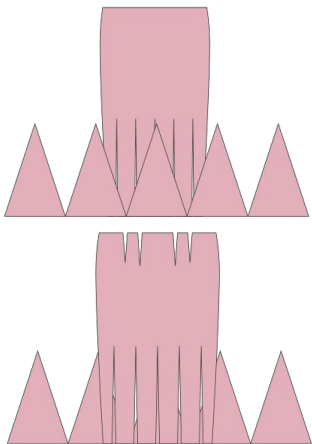
Front & Back Renderings



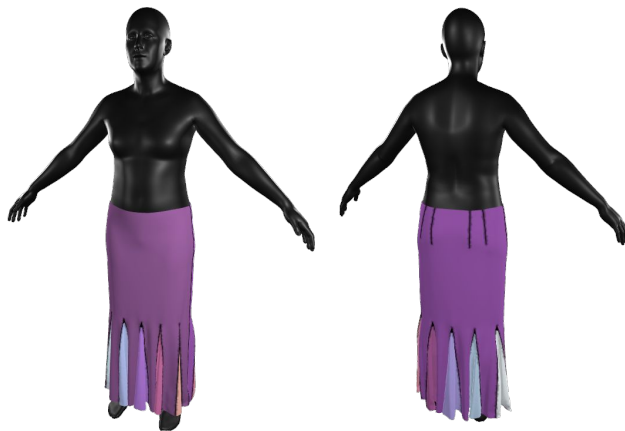
# Annotation Curation

- GarmentCodeData contains large-scale sewing patterns. But they **are not** paired with multimodal annotations.

Sewing Pattern



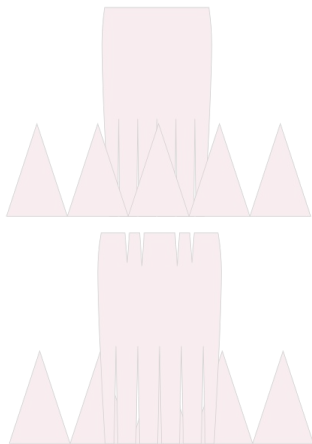
Front & Back Renderings



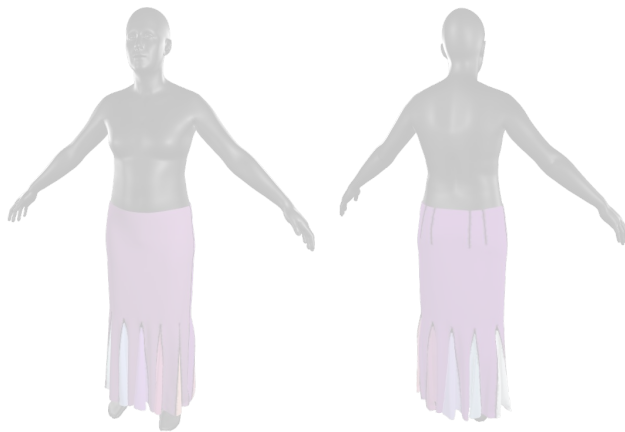
# Annotation Curation

- GarmentCodeData contains large-scale sewing patterns. But they **are not** paired with multimodal annotations.
- We collect multimodal annotations for the sewing patterns:
  - a. **Two types of language descriptions** using GPT4o (Descriptive & Speculative).

Sewing Pattern



Front & Back Renderings



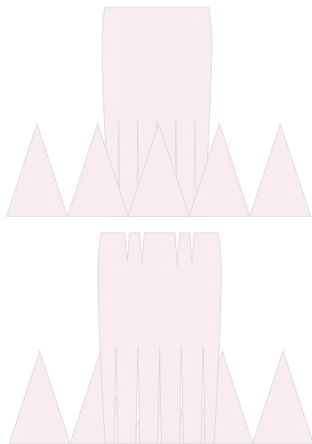
## Descriptive Text

This **maxi godet skirt** features a **knee-length silhouette with 10 inserts**, creating a **flared design from the knee downward**. The skirt's structure is formed by connecting multiple panels at seams spaced evenly around the circumference, with the inserts creating additional volume and flow. **Each insert is integrated between the vertical seams, enhancing the garment's overall shape and movement**. The stitching highlights the transitions between the main panels and the godet inserts, contributing to the skirt's distinctive and fluid appearance.

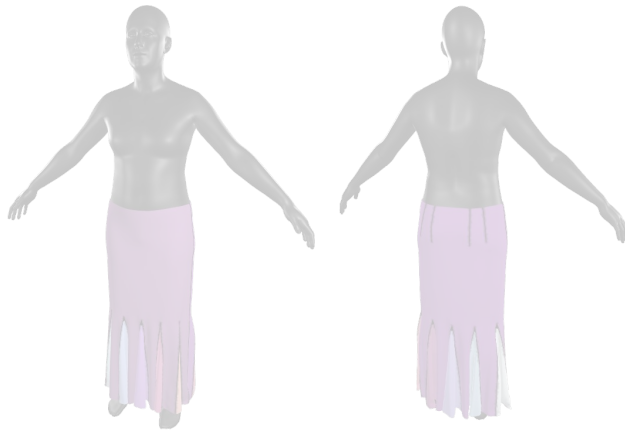
# Annotation Curation

- GarmentCodeData contains large-scale sewing patterns. But they **are not** paired with multimodal annotations.
- We collect multimodal annotations for the sewing patterns:
  - a. **Two types of language descriptions** using GPT4o (Descriptive & Speculative)

Sewing Pattern



Front & Back Renderings



## Descriptive Text

This **maxi godet skirt** features a **knee-length silhouette with 10 inserts**, creating a **flared design from the knee downward**. The skirt's structure is formed by connecting multiple panels at seams spaced evenly around the circumference, with the inserts creating additional volume and flow. **Each insert is integrated between the vertical seams, enhancing the garment's overall shape and movement.** The stitching highlights the transitions between the main panels and the godet inserts, contributing to the skirt's distinctive and fluid appearance.

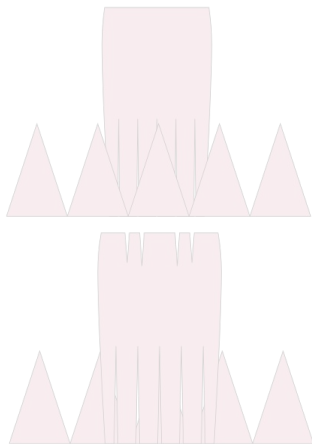
## Speculative Text

This skirt would be appropriate to wear at a **semi-formal evening event**, such as **a garden party** or **an outdoor dinner**.

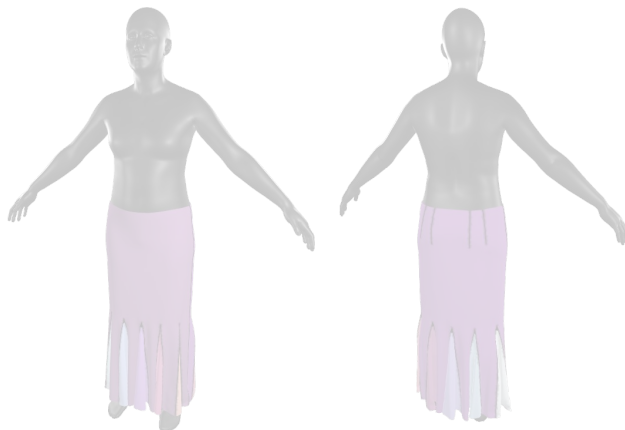
# Annotation Curation

- GarmentCodeData contains large-scale sewing patterns. But they **are not** paired with multimodal annotations.
- We collect multimodal annotations for the sewing patterns:
  - a. **Two types of language descriptions** using GPT4o (Descriptive & Speculative).
  - b. Paired Sewing pattern and instruction for **editing**.

Sewing Pattern



Front & Back Renderings



## Texts

This maxi godet skirt features a knee-length silhouette with 10 inserts, creating a flared design from the knee downward. The skirt's structure is formed by connecting multiple panels at seams spaced evenly around the circumference, with the inserts creating additional volume and flow. Each insert is integrated between the vertical seams, enhancing the garment's overall shape and movement. The stitching highlights the transitions between the main panels and the godet inserts, contributing to the skirt's distinctive and fluid appearance.

This skirt would be appropriate to wear at a semi-formal evening event, such as a garden party or an outdoor dinner.

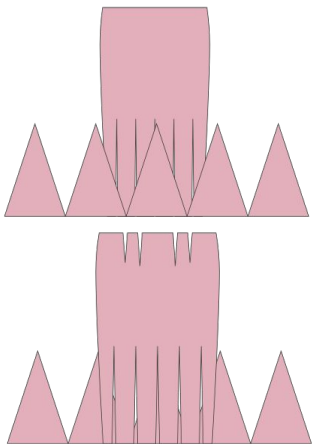
A godet skirt with more inserts



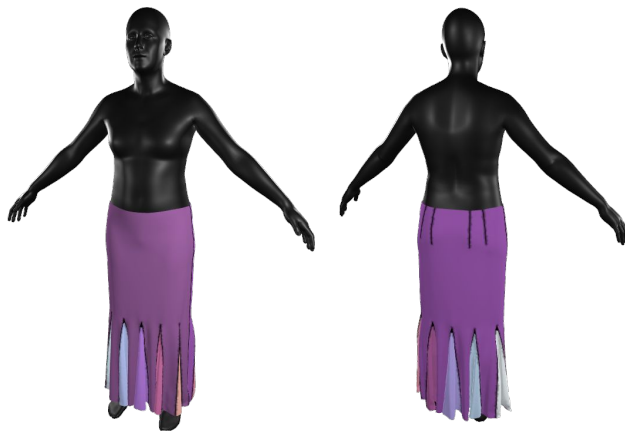
# Annotation Curation

- GarmentCodeData contains large-scale sewing patterns. But they **are not** paired with multimodal annotations.
- We collect multimodal annotations for the sewing patterns:
  - a. **Two types of language descriptions** using GPT4o (Descriptive & Speculative).
  - b. Paired Sewing pattern and instruction for **editing**.

Sewing Pattern



Front & Back Renderings



## Texts

This **maxi godet skirt** features a **knee-length silhouette with 10 inserts**, creating a **flared design from the knee downward**. The skirt's structure is formed by connecting multiple panels at seams spaced evenly around the circumference, with the inserts creating additional volume and flow. **Each insert is integrated between the vertical seams, enhancing the garment's overall shape and movement.** The stitching highlights the transitions between the main panels and the godet inserts, contributing to the skirt's distinctive and fluid appearance.

This skirt would be appropriate to wear at a **semi-formal evening event** such as a **garden party** or an **outdoor dinner**.

A godet skirt with more inserts

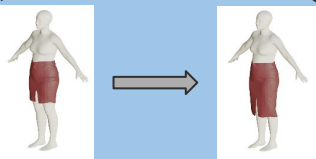


# Road Map

## Data:

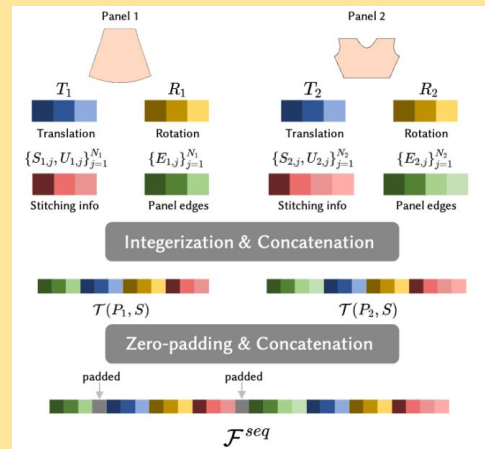
How do we build a large scale multimodal dataset of sewing patterns?

It's my best friend's wedding! What should I wear?



## Representation:

How do we represent sewing pattern for MLLMs?

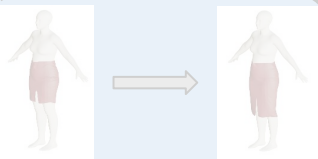
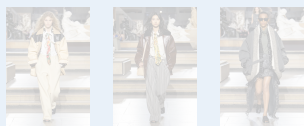


# Road Map

## Data:

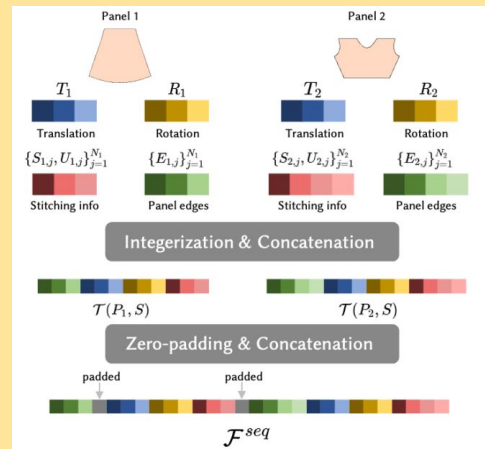
How do we build a large scale multimodal dataset of sewing patterns?

It's my best friend's wedding! What should I wear?



## Representation:

How do we represent sewing pattern for MLLMs?



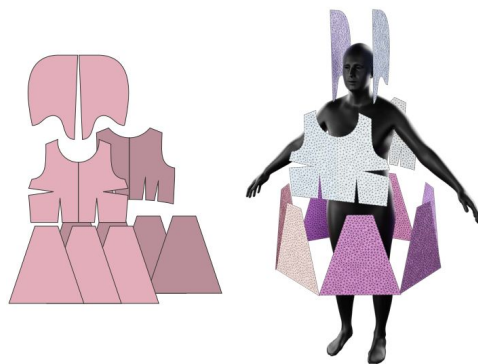
# Reminder

For us, sewing patterns are a set of 2D panels  $P_i$ ,  $i = 1, \dots, n$ .

Each  $P_i$  consists of

1. (optionally) A name,
2. A 3D rigid transformation  $R_i$  &  $T_i$ ,
3. An ordered set of edges  $(e_i^1, e_i^2, \dots, e_i^{m_i})$ , that are either straight lines, cubic bezier curves, or arcs.
4. Vertices  $\{v_i^s\}_{s=1, \dots, m_i}$  and control points  $\{c_i^k\}_{k=1, \dots, m_i}$ .

A set of edge pairs  $\{e_i^{k_i}, e_j^{l_j}\}$  indicates if two edges are stitched together (both within and across panels).



# Reminder

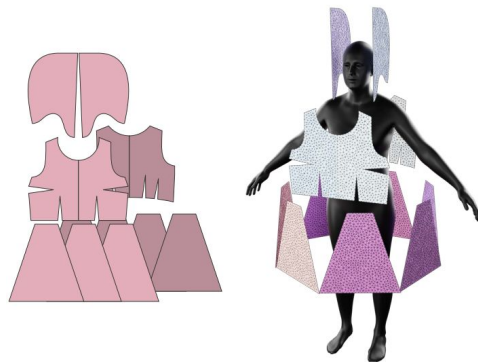
For us, sewing patterns are a set of 2D panels  $P_i$ ,  $i = 1, \dots, n$ .

Each  $P_i$  consists of

1. (optionally) A name,
2. A 3D rigid transformation  $R_i$  &  $T_i$ ,
3. An ordered set of edges  $(e_i^1, e_i^2, \dots, e_i^{m_i})$ , that are either straight lines, cubic bezier curves, or arcs.
4. Vertices  $\{v_i^s\}_{s=1, \dots, m_i}$  and control points  $\{c_i^k\}_{k=1, \dots, m_i}$ .

Contains both  
continuous and  
discrete parameters.

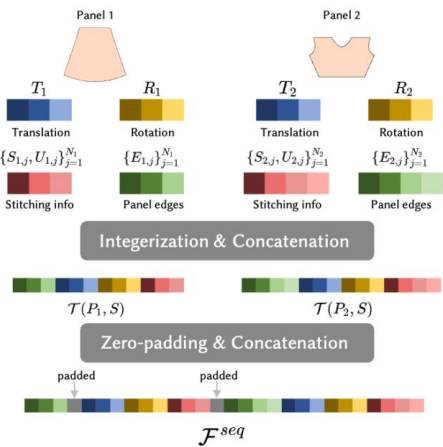
A set of edge pairs  $\{e_i^{k_i}, e_j^{l_j}\}$  indicates if two edges are stitched together (both within and across panels).



# Sewing Pattern Tokenization

- Sewing patterns consists of **discrete** and **continuous** parameters.
  - **Discrete**: Variable number of panels ( $P_i$ ), edges ( $e_i^l$ ), and stitches ( $\{e_i^{k_i}, e_j^{l_j}\}$ ).
  - **Continuous**: 3D transformations ( $[R_i|T]_i$ ) and edge parameters ( $\{v_i^s\}$  and  $\{c_i^k\}$ ).

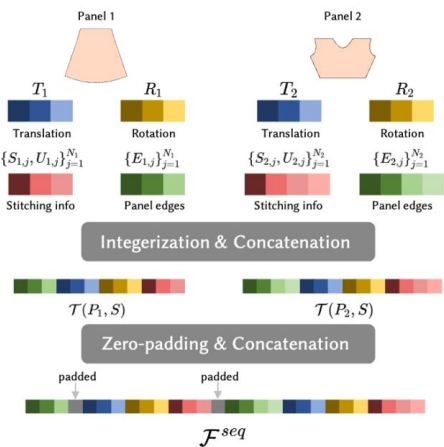
Previous Method (DressCode):  
pad everything to fixed length.



# Sewing Pattern Tokenization

- Sewing patterns consists of **discrete** and **continuous** parameters.
  - **Discrete**: Variable number of panels ( $P_i$ ), edges ( $e_i^l$ ), and stitches ( $\{e_i^{k_i}, e_j^{l_j}\}$ ).
  - **Continuous**: 3D transformations ( $[R_i|T]_i$ ) and edge parameters ( $\{v_i^s\}$  and  $\{c_i^k\}$ ).

Previous Method (DressCode):  
**pad everything to fixed length.**



Problem: **context inefficient** and **impractical** for complex garments.

E.g., each pattern needs **> 30k tokens** for patterns in GamentCodeData.

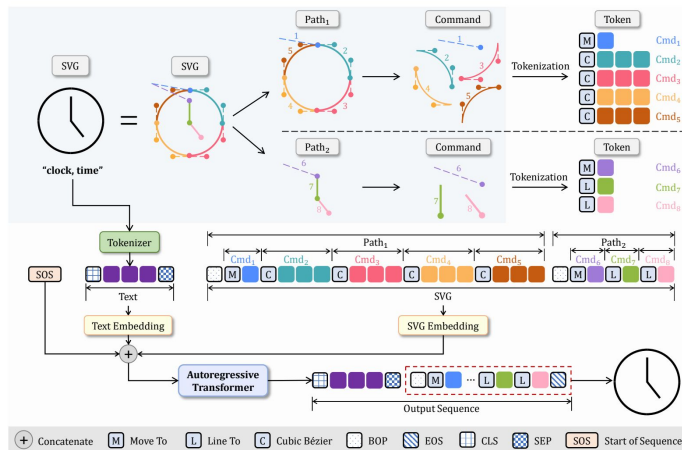
# Sewing Pattern Tokenization

- Sewing patterns consists of **discrete** and **continuous** parameters.
  - **Discrete**: Variable number of panels ( $P_i$ ), edges ( $e_i^l$ ), and stitches ( $\{e_i^{k,i}, e_j^{l,j}\}$ ).
  - **Continuous**: 3D transformations ( $[R_i|T_i]$ ) and edge parameters ( $\{v_i^s\}$  and  $\{c_i^k\}$ ).

Previous Method (DressCode):  
pad everything to fixed length.



Inspired by SVG generation methods, we represent sewing patterns as a set of drawing commands.



# Our Tokenization Scheme

4 types of edges: **LINE**, **QCURVE**, **CCURVE**, **ARC**.



# Our Tokenization Scheme

4 types of edges: LINE, QCURVE, CCURVE, ARC.



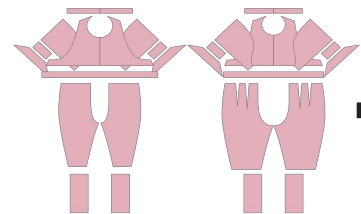
4 special tokens: <Start Panel>, <End Panel>, <Start Pattern>, <End Pattern>.

SP

EP

SG

EG



# Our Tokenization Scheme

4 types of edges: LINE, QCURVE, CCURVE, ARC.



4 special tokens: <Start Panel>, <End Panel>, <Start Pattern>, <End Pattern>.

SP

EP

SG

EG

Each edge associates with a stitch tag  $\langle \text{tag}_1, \dots, \text{tag}_M, \text{tag}_{\text{null}} \rangle$ .

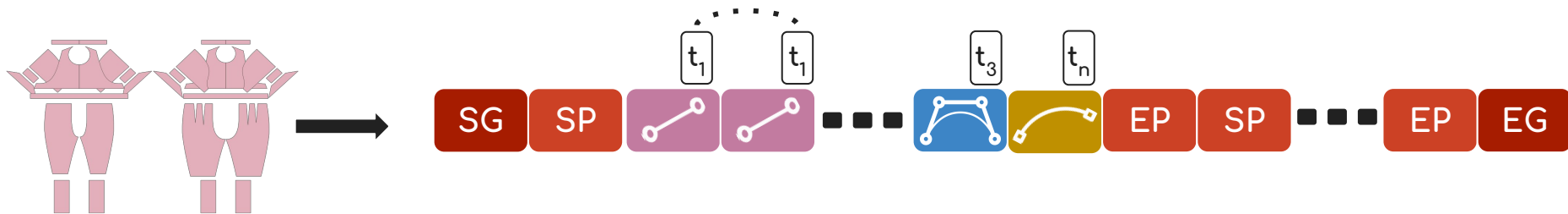
- Two edges has the same tag  $\Leftrightarrow$  Stitched together.
- Null tag = no stitching.

$t_1$

...

$t_M$

$t_n$



# Our Tokenization Scheme

4 types of edges: **LINE**, **QCURVE**, **CCURVE**, **ARC**.



4 special tokens: **<Start Panel>**, **<End Panel>**, **<Start Pattern>**, **<End Pattern>**.

SP

EP

SG

EG

Each edge associates with a stitch tag  $\langle \text{tag}_1, \dots, \text{tag}_M, \text{tag}_{\text{null}} \rangle$ .

→ Two edges has the same tag  $\Leftrightarrow$  Stitched together.

→ Null tag = no stitching.

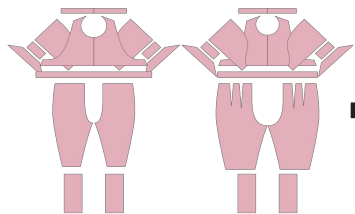
$t_1$

...

$t_M$

$t_n$

Vertices and control points are **regressed with regression heads**.



# Our Tokenization Scheme

4 types of edges: LINE, QCURVE, CCURVE, ARC.



4 special tokens: <Start Panel>, <End Panel>, <Start Pattern>, <End Pattern>.

SP

EP

SG

EG

Each edge associates with a stitch tag  $\langle \text{tag}_1, \dots, \text{tag}_M, \text{tag}_{\text{null}} \rangle$ .

→ Two edges has the same tag  $\Leftrightarrow$  Stitched together.

→ Null tag = no stitching.

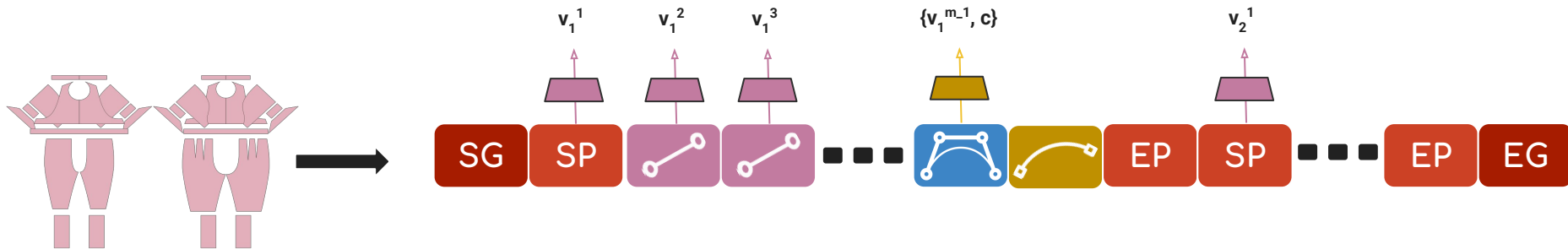
$t_1$

...

$t_M$

$t_n$

Vertices and control points are regressed with regression heads.



# Our Tokenization Scheme

4 types of edges: **LINE**, **QCURVE**, **CCURVE**, **ARC**.



4 special tokens: **<Start Panel>**, **<End Panel>**, **<Start Pattern>**, **<End Pattern>**.

SP

EP

SG

EG

Each edge associates with a stitch tag  $\langle \text{tag}_1, \dots, \text{tag}_M, \text{tag}_{\text{null}} \rangle$ .

- Two edges has the same tag  $\Leftrightarrow$  Stitched together.
- Null tag = no stitching.

$t_1$

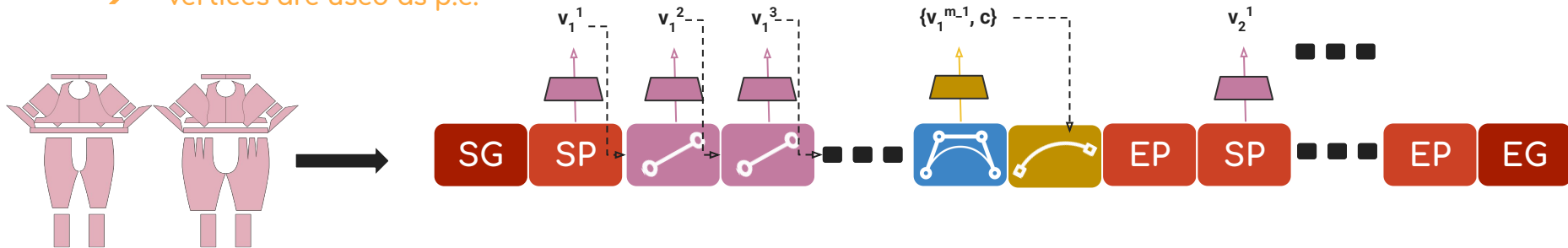
...

$t_M$

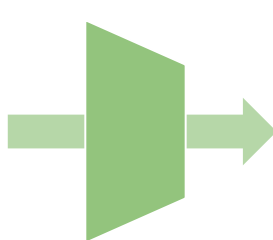
$t_n$

Vertices and control points are **regressed with regression heads**.

- Vertices are used as p.e.



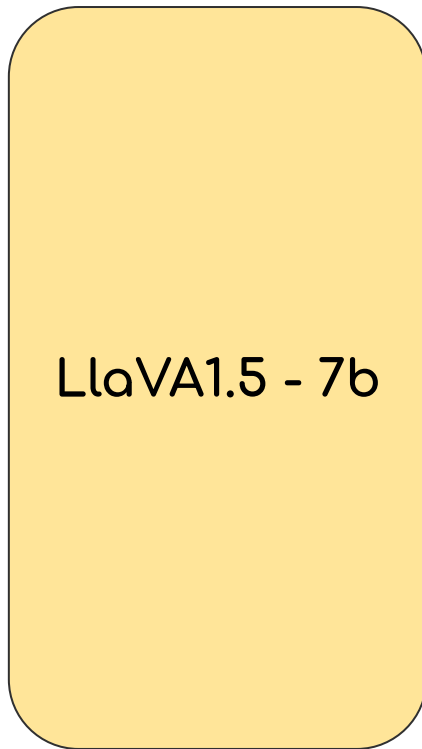
# Training



This skirt would be appropriate to wear at a semi-formal evening event, such as a garden party or an outdoor dinner.



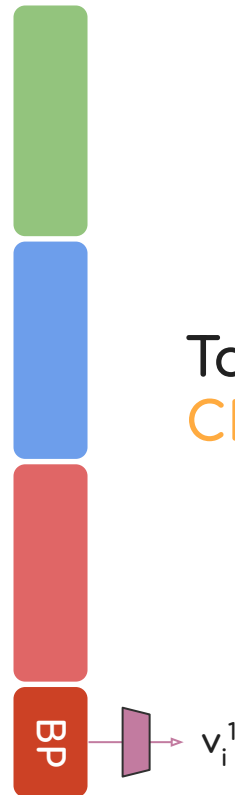
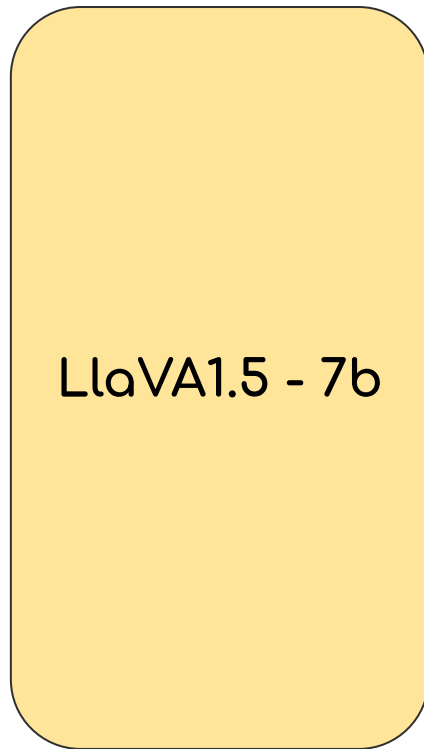
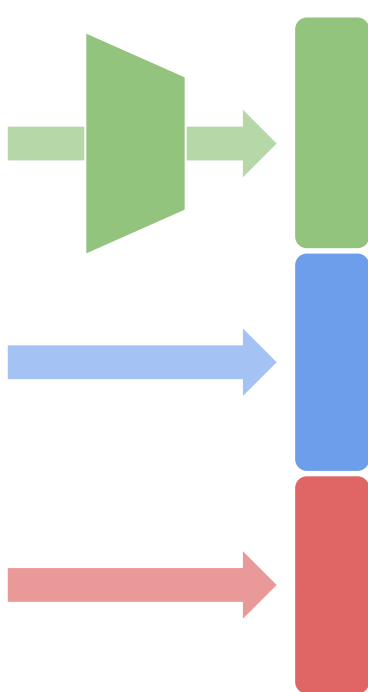
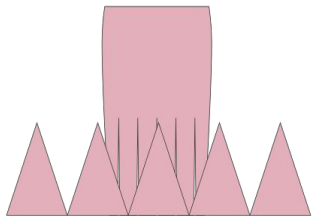
LlaVA1.5 - 7b



# Training



This skirt would be appropriate to wear at a semi-formal evening event, such as a garden party or an outdoor dinner.



Total Loss =  
 $CE + L_2$

# Results -- Single Modal Generation

Our model can generate complex sewing patterns from image inputs.

The previous SoTA fails to converge because of the complexity.

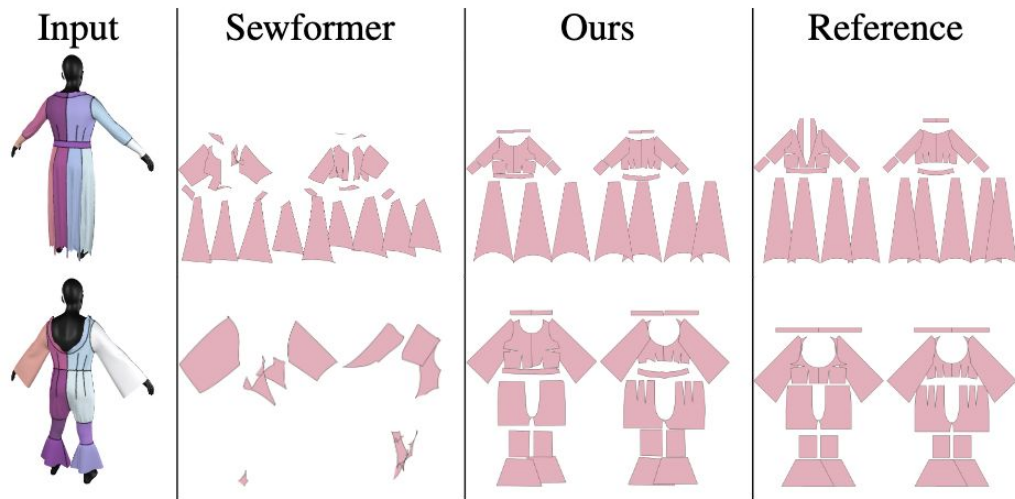


Figure 5. Image to garment.

# Results -- Single Modal Generation

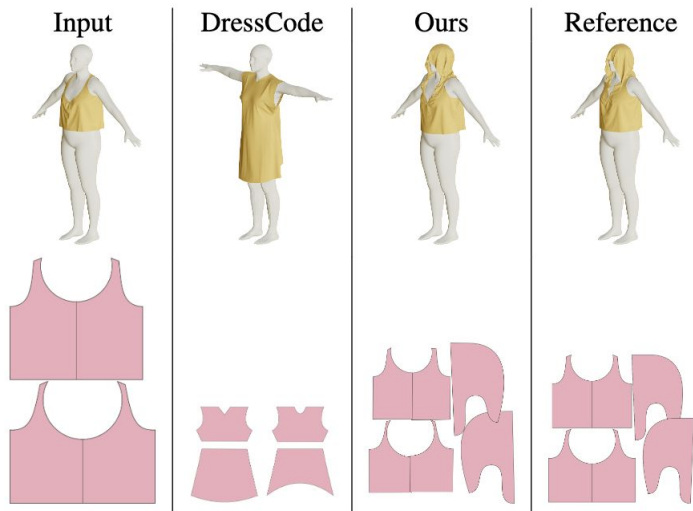
We are also better than the baseline when trained on the previous dataset.

Data	Method	Panel L2 (↓)	#Panel Acc (↑)	Edge Acc (↑)	Rot L2 (↓)	Transl L2 (↓)	Stitch Acc (↑)
Sewfactory	SewFormer	3.34	89.78	99.34	0.0078	0.82	99.17
	AIpparel	<b>2.77</b>	<b>93.90</b>	<b>99.95</b>	<b>0.0051</b>	<b>0.60</b>	<b>99.75</b>
GarmentCode	SewFormer-FT	12.27	79.40	44.70	0.037	4.47	2.78
	AIpparel	<b>5.39</b>	<b>85.20</b>	<b>82.70</b>	<b>0.015</b>	<b>2.72</b>	<b>77.2</b>

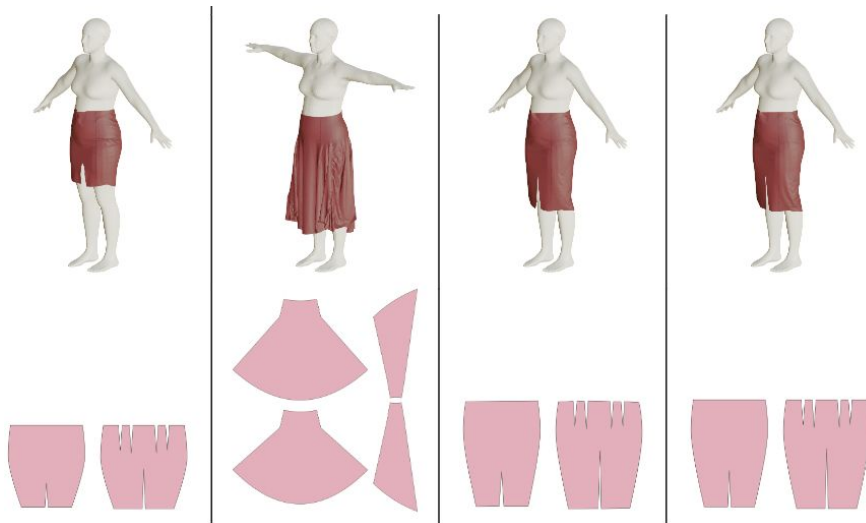
# Results -- Editing

We can also edit sewing patterns based on language instructions.

Include a hood  
in the design.



Make the skirt  
longer.



# Ablation on Our Tokenization

To validate our tokenization, we fix the backbone and replace the sewing pattern tokenizer with the previous work.

**Result:** we are both qualitative better and faster, since the previous tokenization scheme generates lots of zero paddings.

Methods	Accuracy (↑)	Shape L2 (↓)	Time (↓)
DressCode	38.4	22.4	52.2
Ours w.o. reg.	79.0	7.17	3.43
Ours	<b>85.0</b>	<b>6.05</b>	<b>2.08</b>