Evolutionary Search for Fashion Styles in the Latent Space of Generative Adversarial Networks

A thesis presented to the Faculty of Humanities in partial fulfillment of the requirements for the degree

Master of Science in IT & Cognition

June 28th, 2021

Imke Grabe

Supervisor: Manex Agirrezabal

KØBENHAVNS UNIVERSITET

Evolutionary Search for Fashion Styles in the Latent Space of Generative Adversarial Networks

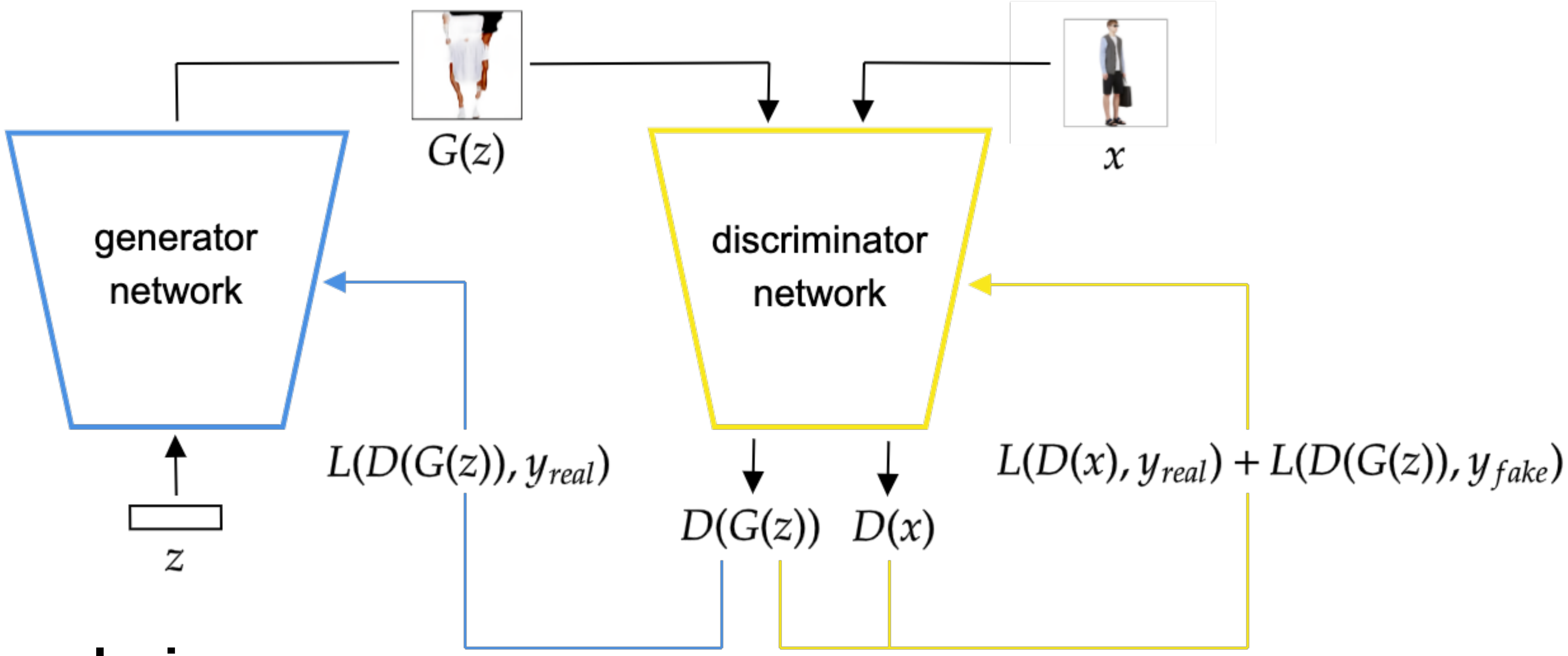
- **Introduction and Background**
- **Definition of the Task**
- **Method**
- **Results and Discussion**
- **Conclusion**

Evolutionary Search for Fashion Styles in the Latent Space of Generative Adversarial Networks

Purpose of the study: Investigate the application of a genetic algorithm to find latent vectors in the latent space of Generative Adversarial Networks, that represent certain fashion styles, previously discovered by a clustering model.

Why Generative Adversarial Networks?

- **Generative Adversarial Nets** (GANs) by Goodfellow et al. (2014)
- Generative deep learning



Control of generative clothing design:

- encoding of a text description in the latent vector (Zhu et al., 2017)
- disentangling color, texture, and shape (Yildirim et al., 2018)

Why Fashion Styles?

Fashion is a “a specific category of clothes (…) determine[d] by a process of style change” (Mackinney-Valentin, 2010, p.19).

visual themes

(Matzen et al., 2017, p.1)

spatial and temporal relations

(Al-Halah and Grauman, 2020, p.12,15)

Why Fashion Styles?

Fashion is a “a specific category of clothes (…) determine[d] by a process of style change” (Mackinney-Valentin, 2010, p.19).

visual themes

(Matzen et al., 2017, p.1)

spatial and temporal relations

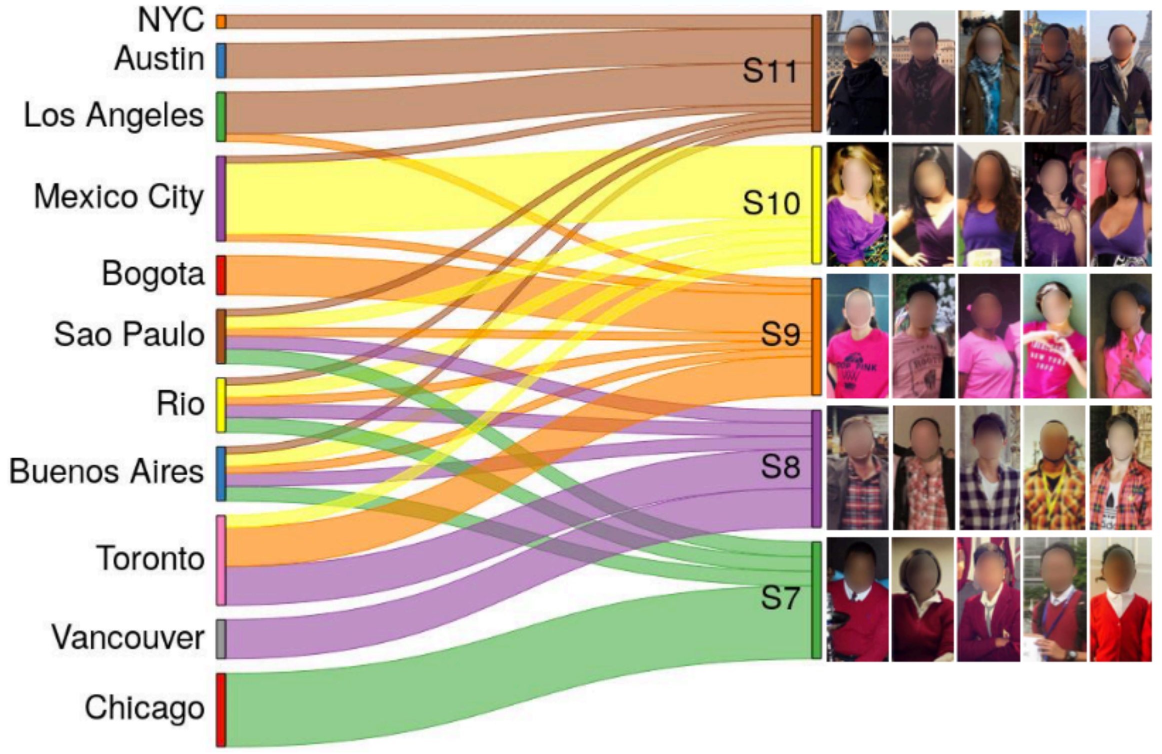
(Al-Halah and Grauman, 2020, p.12,15)

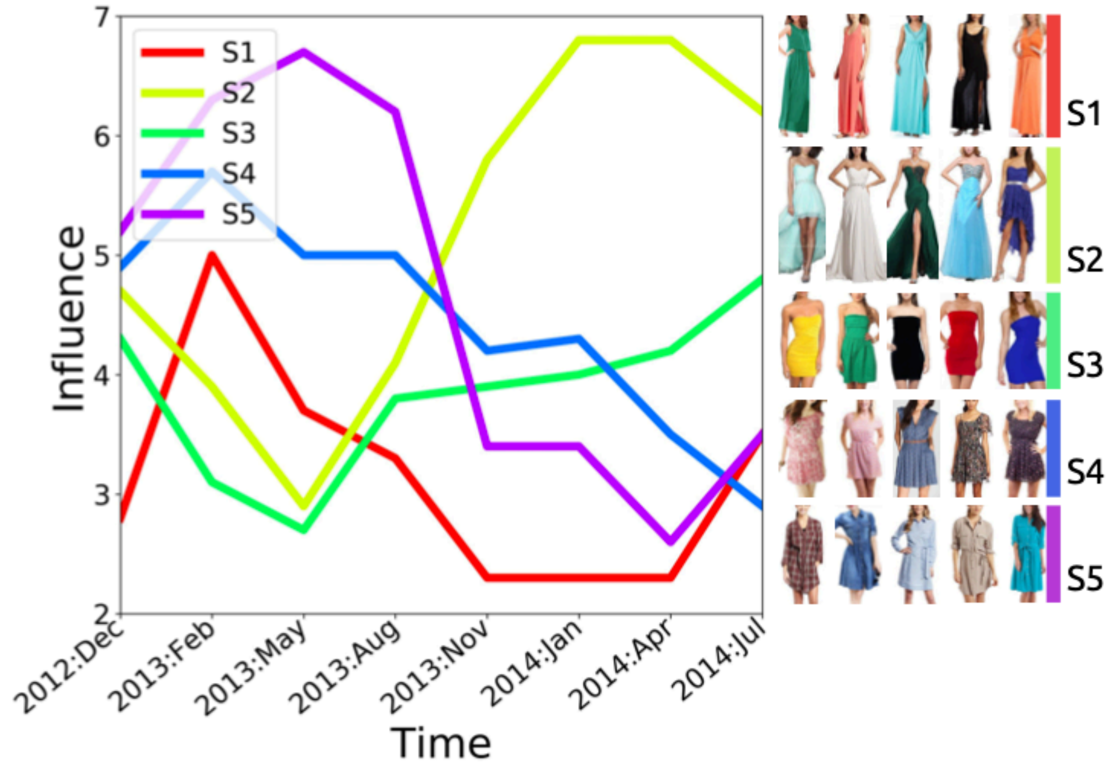
Why Fashion Styles?

Fashion is a “a specific category of clothes (…) determine[d] by a process of style change” (Mackinney-Valentin, 2010, p.19).

visual themes

(Matzen et al., 2017, p.1)





spatial and temporal relations

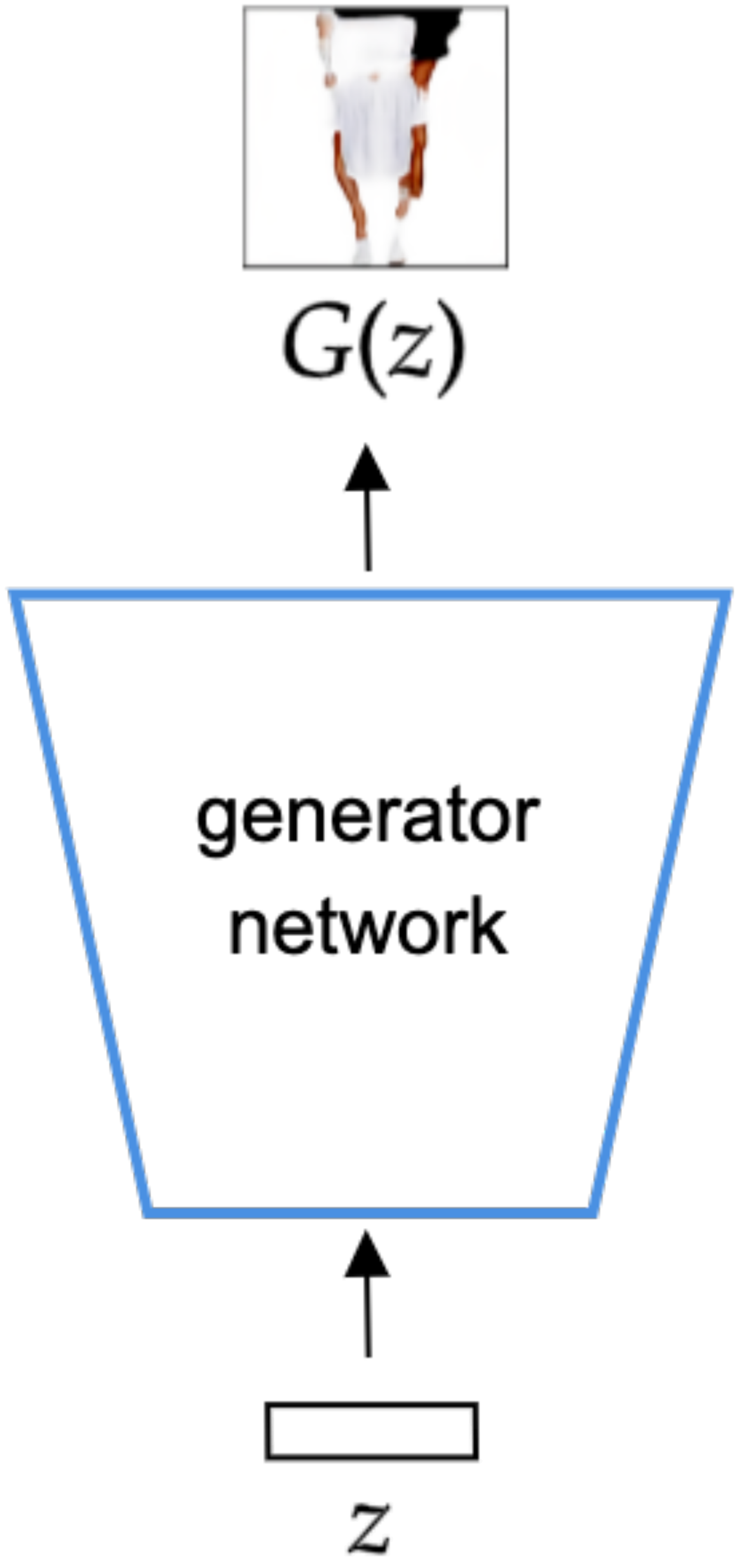
(Al-Halah and Grauman, 2020, p.12,15)

Why Evolutionary Search?

- evolutionary latent space exploration in GANs (Fernandes et al., 2020)
- smoothness of latent space (Bontrager et al, 2018)
- large entangled design space

Fernandes et al., 2020, p.13

Bontrager et al, 2018, p.11

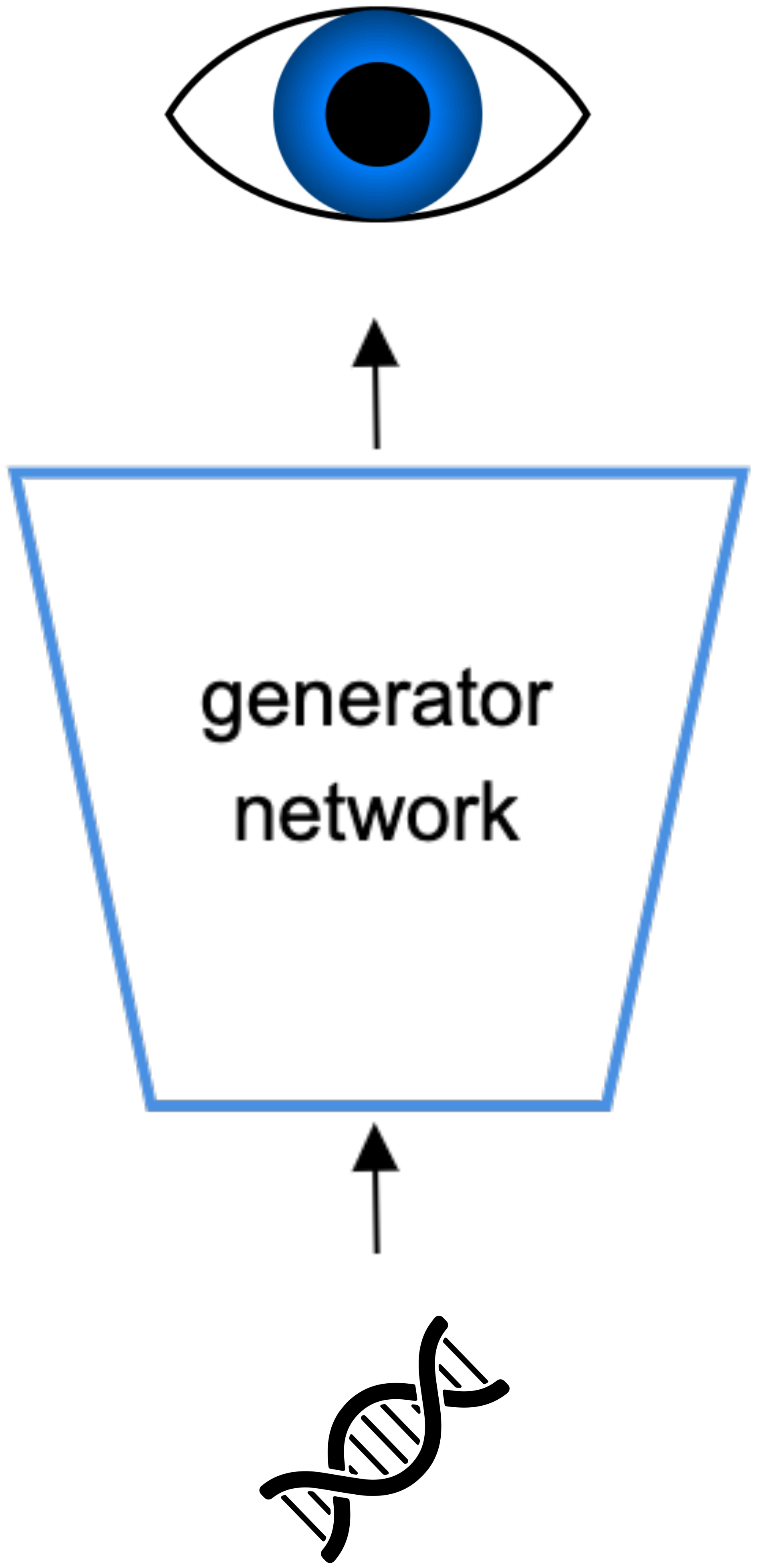


Why Evolutionary Search?

- evolutionary latent space exploration in GANs (Fernandes et al., 2020)
- smoothness of latent space (Bontrager et al, 2018)
- large entangled design space

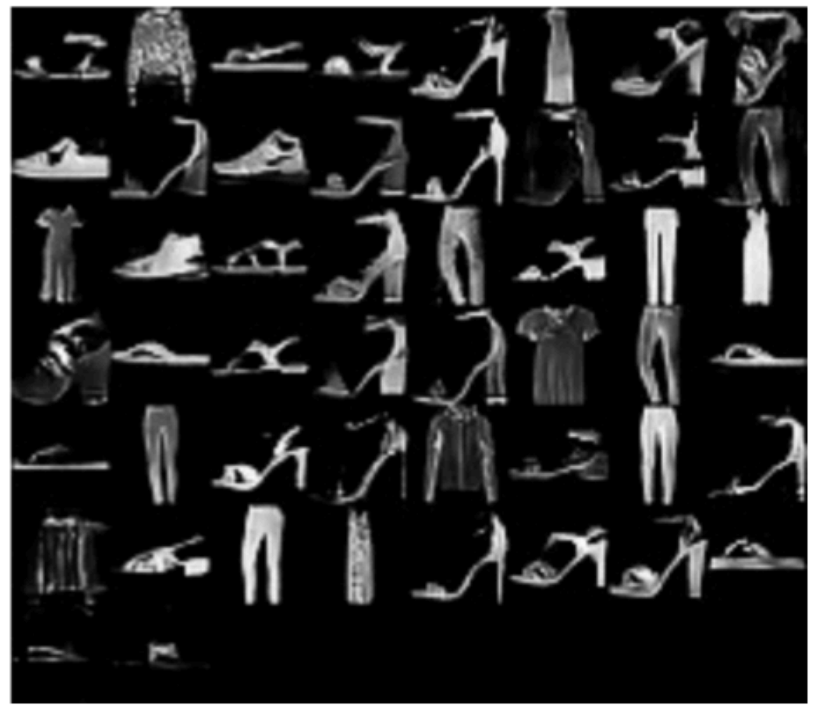
Fernandes et al., 2020, p.13

Bontrager et al, 2018, p.11



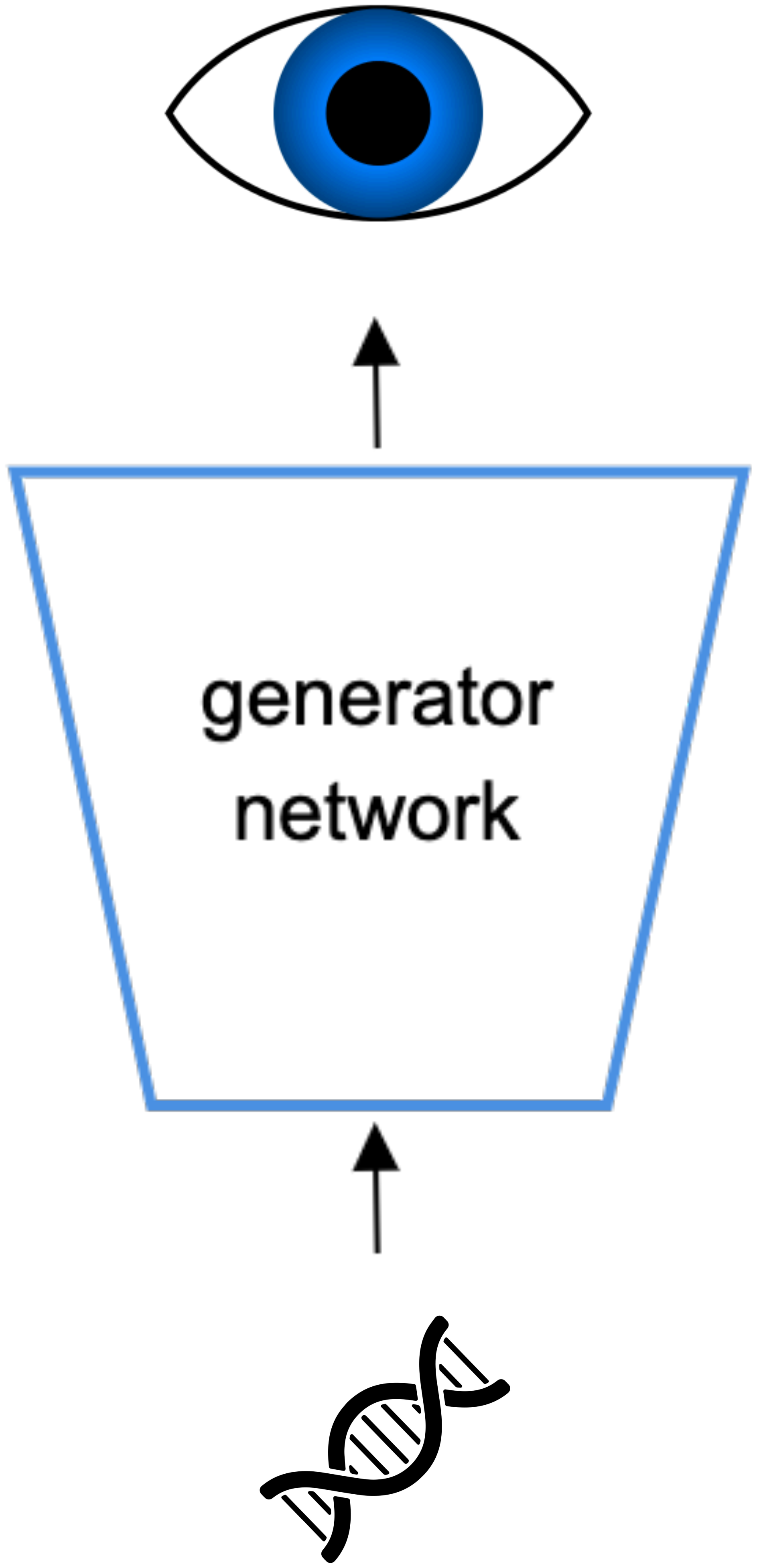
Why Evolutionary Search?

- evolutionary latent space exploration in GANs (Fernandes et al., 2020)
- smoothness of latent space (Bontrager et al, 2018)
- large entangled design space



(Fernandes et al., 2020, p.13)

(Bontrager et al, 2018, p.11)



Research Question

Can a genetic algorithm guide a GAN's generation of designs towards a specific fashion style?

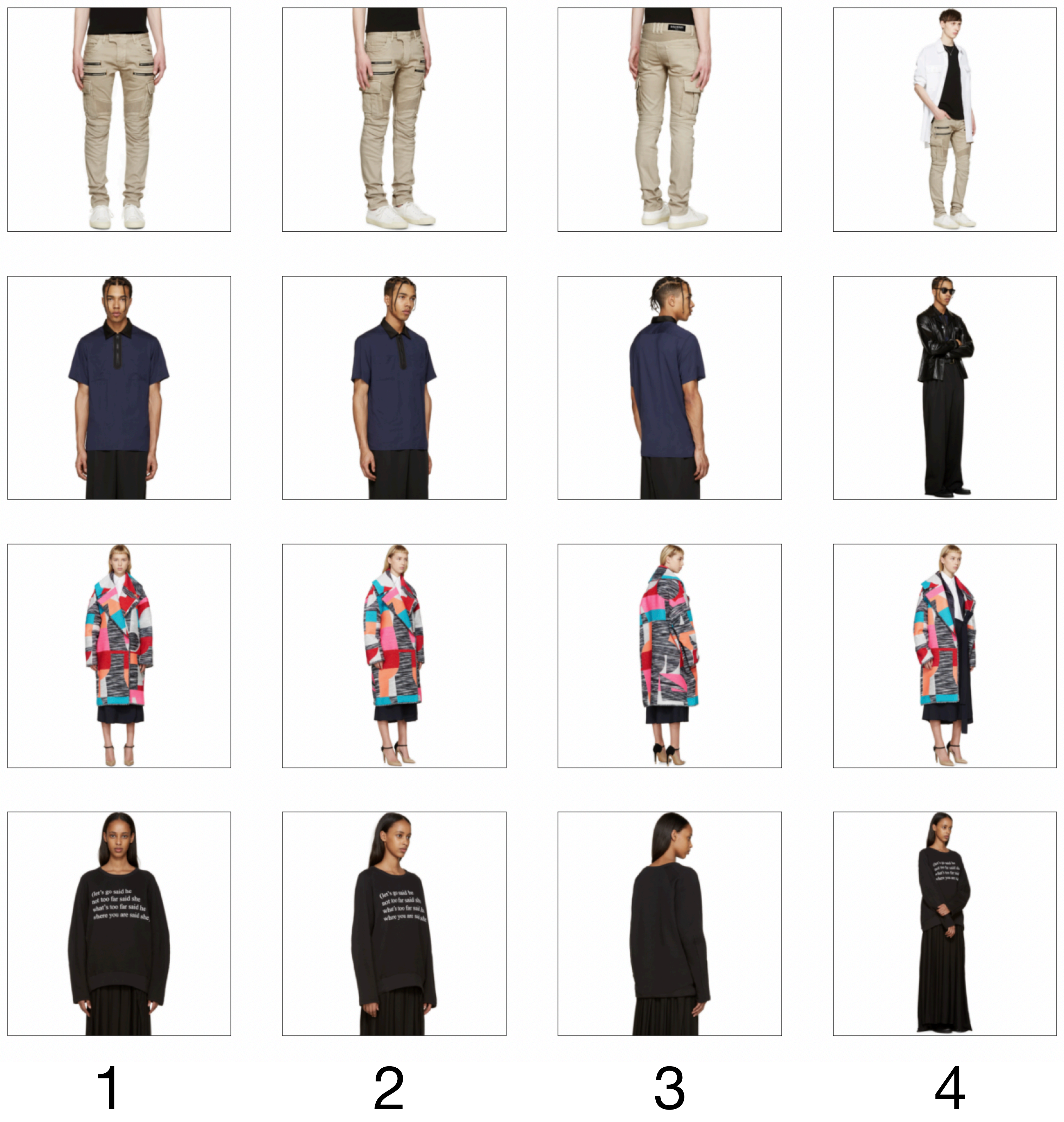
Task

1) Discover fashion style clusters
2) Train a generative model
3) Find styles in the generative model's latent space

Dataset

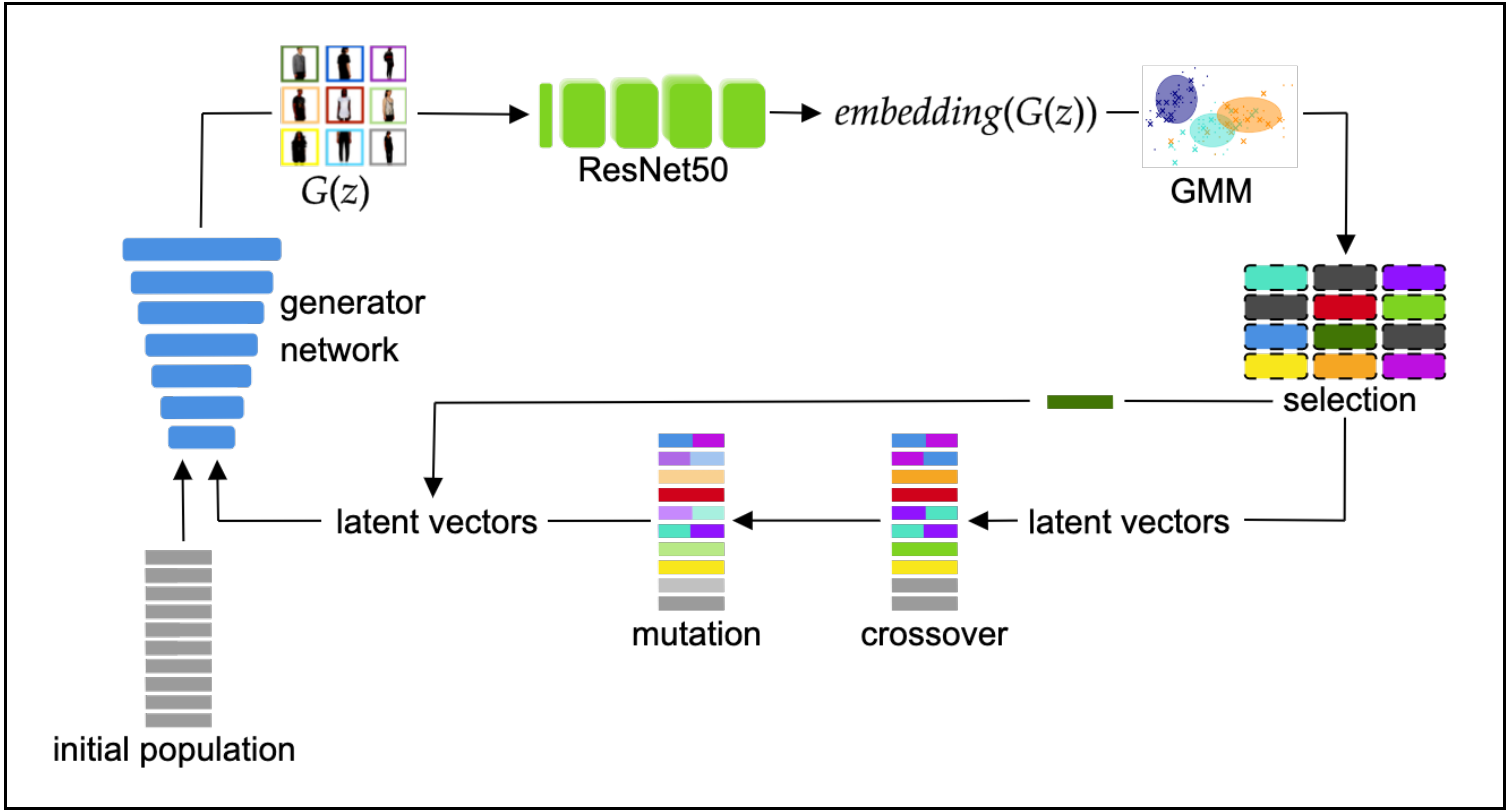
Clothing partition of *FashionGen* (Rostamzadeh et al., 2018)

Pose 1	49,108
Pose 2	49,069
Pose 3	49,061
Pose 4	49,010
Total	196,248



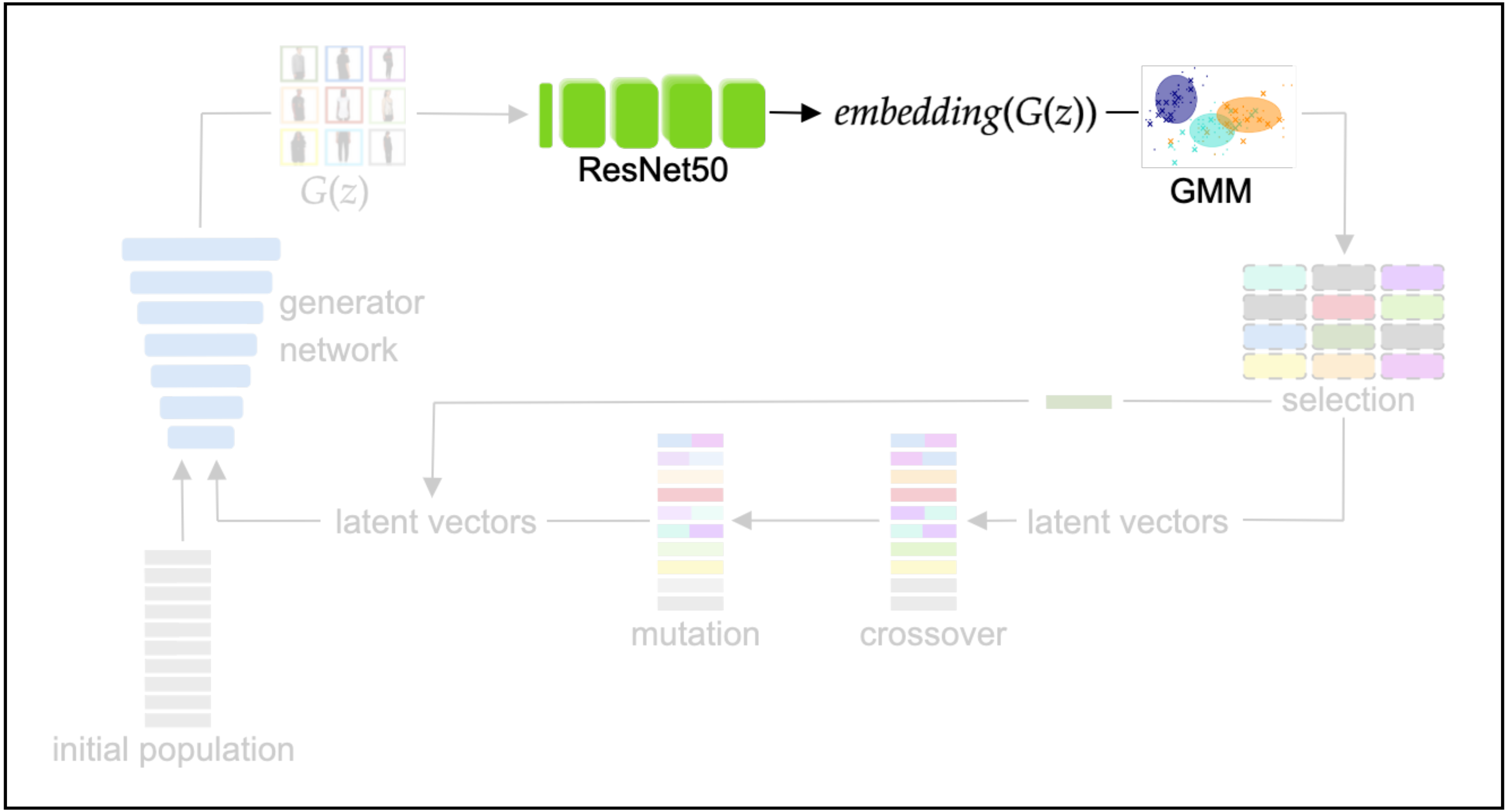
- online webshop SSENSE
- plain white background and consistent lighting conditions
- 196,248 images equally distributed across four poses per outfit
- 256^2 pixels resolution

Model

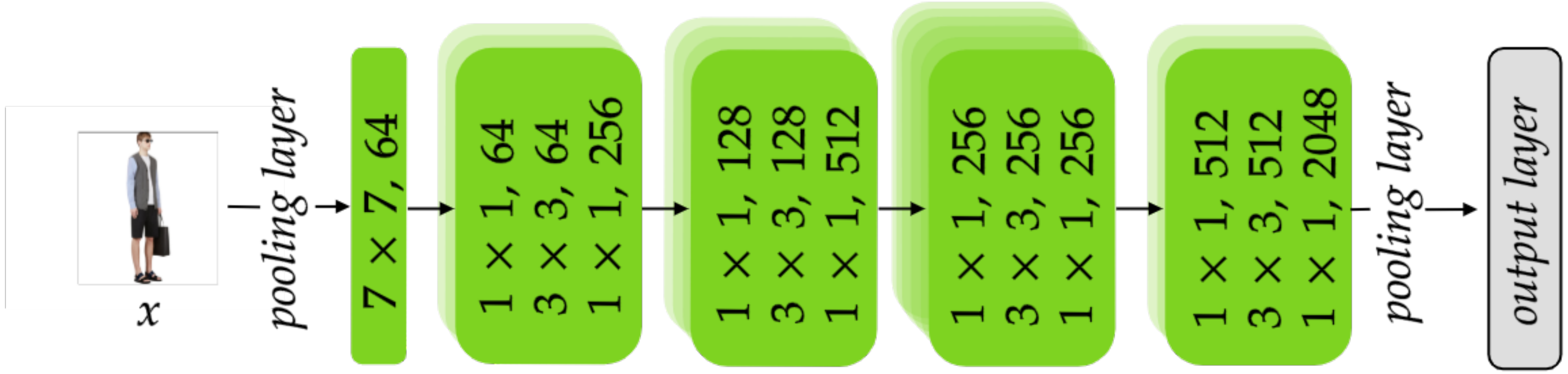


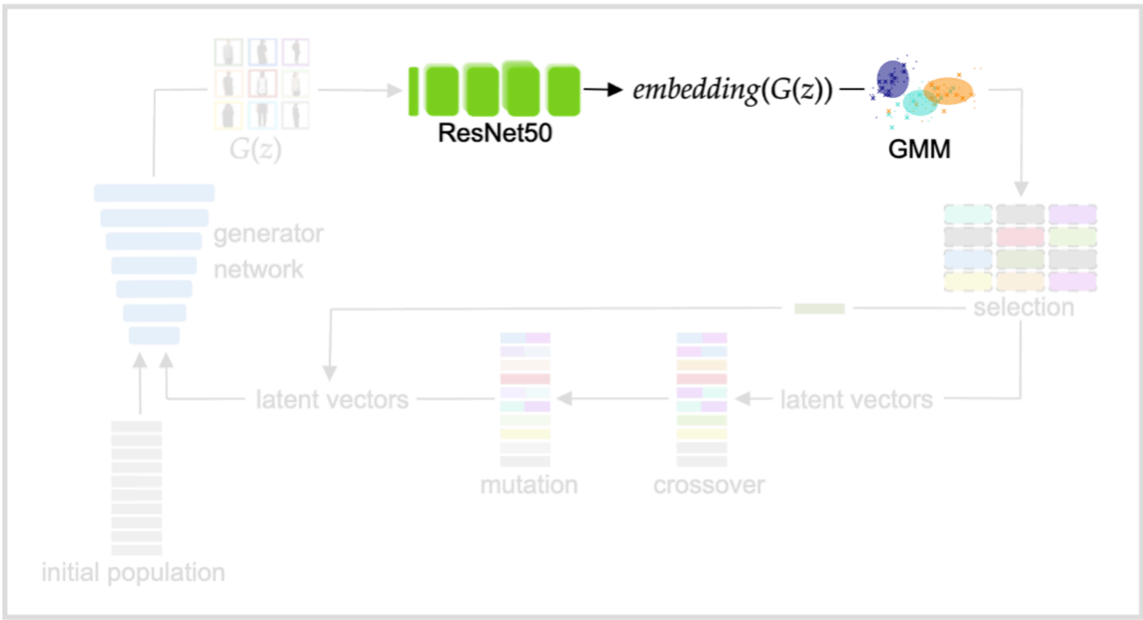
$G(z)$
ResNet50
$embedding(G(z))$
GMM
generator
network
selection
latent vectors
mutation
crossover
latent vectors
initial population

1) Style model



1) Style model

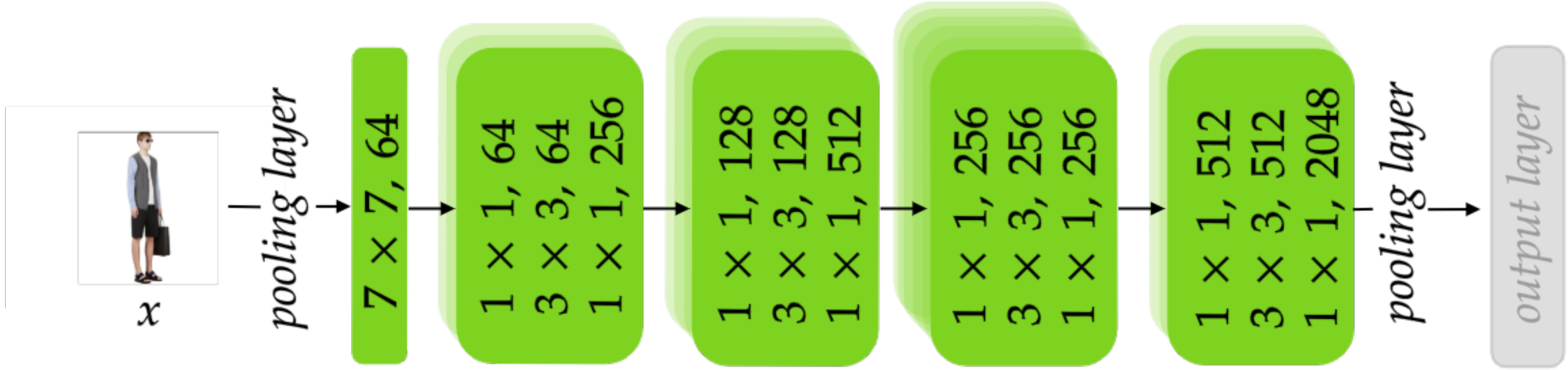


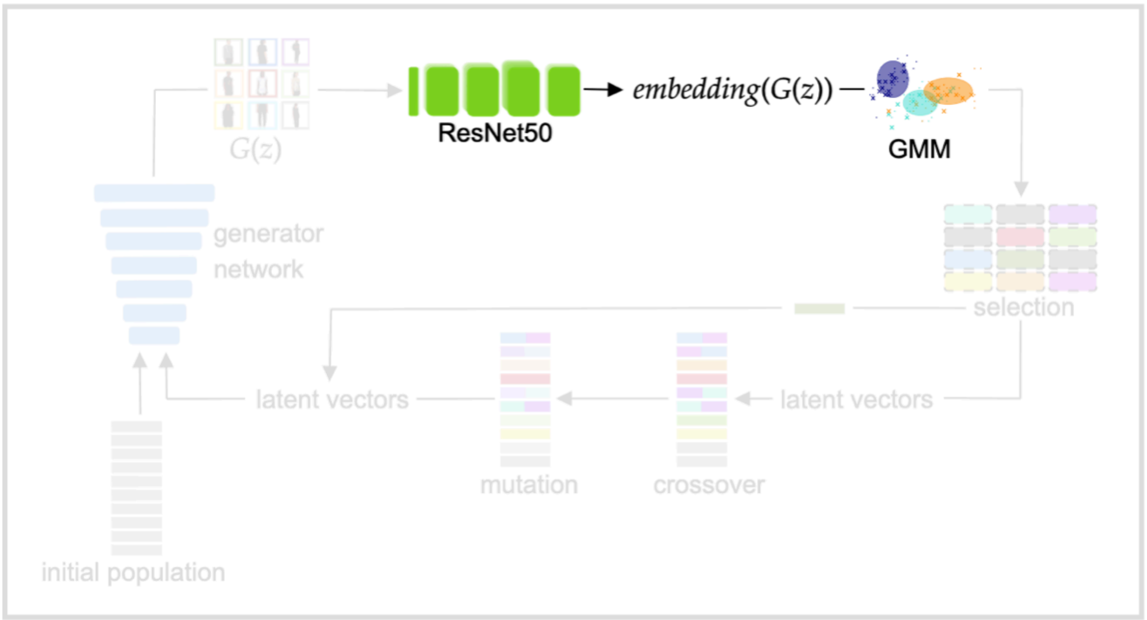


Residual Neural Network (ResNet)

- ResNet50 (Takagi et al., 2017)
- pre-trained on DeepFashion (Al-Halah and Grauman, 2020)

1) Style model

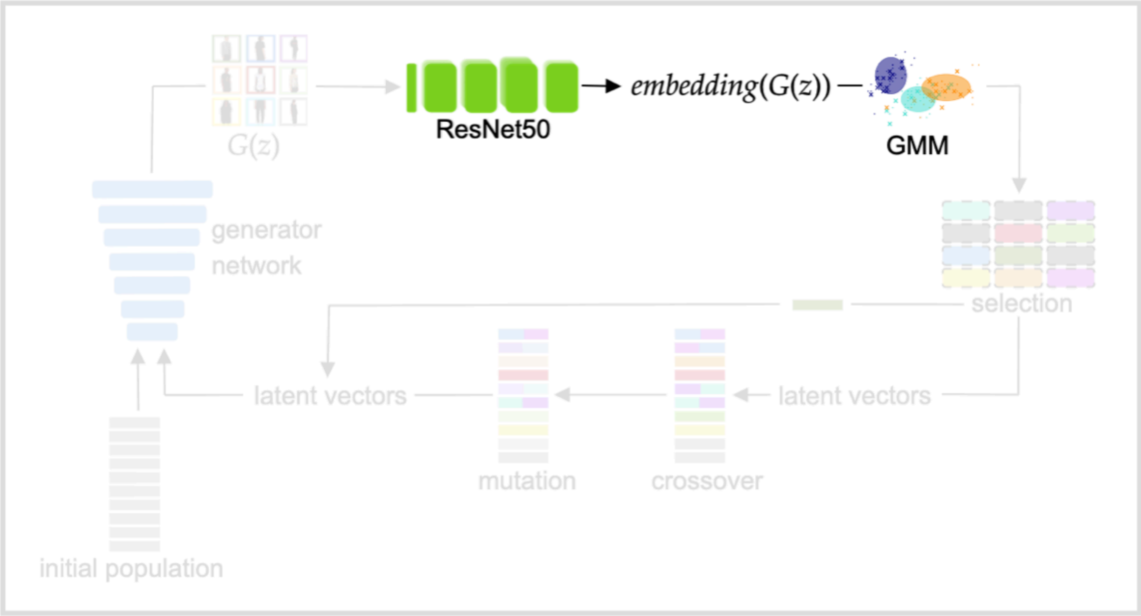


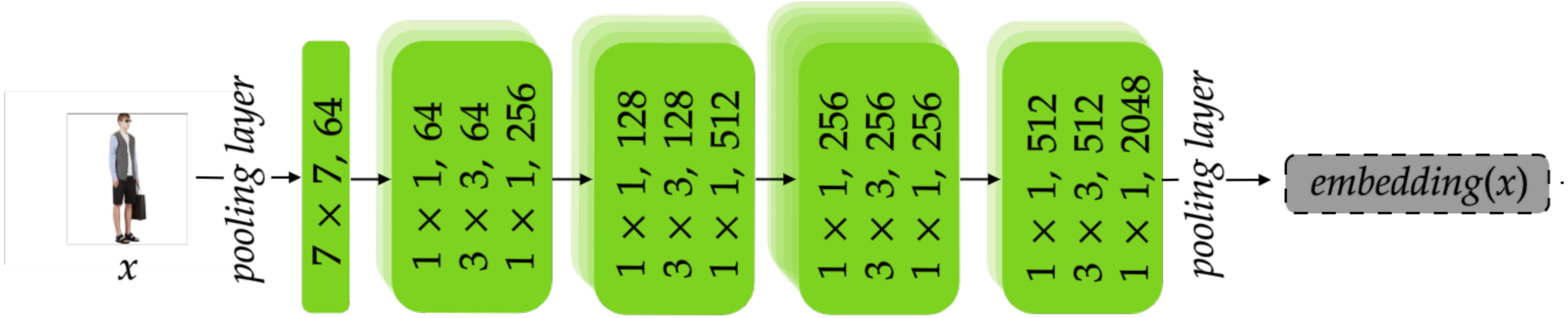


Residual Neural Network (ResNet)

- ResNet50 (Takagi et al., 2017)
- pre-trained on DeepFashion (Al-Halah and Grauman, 2020)

1) Style model

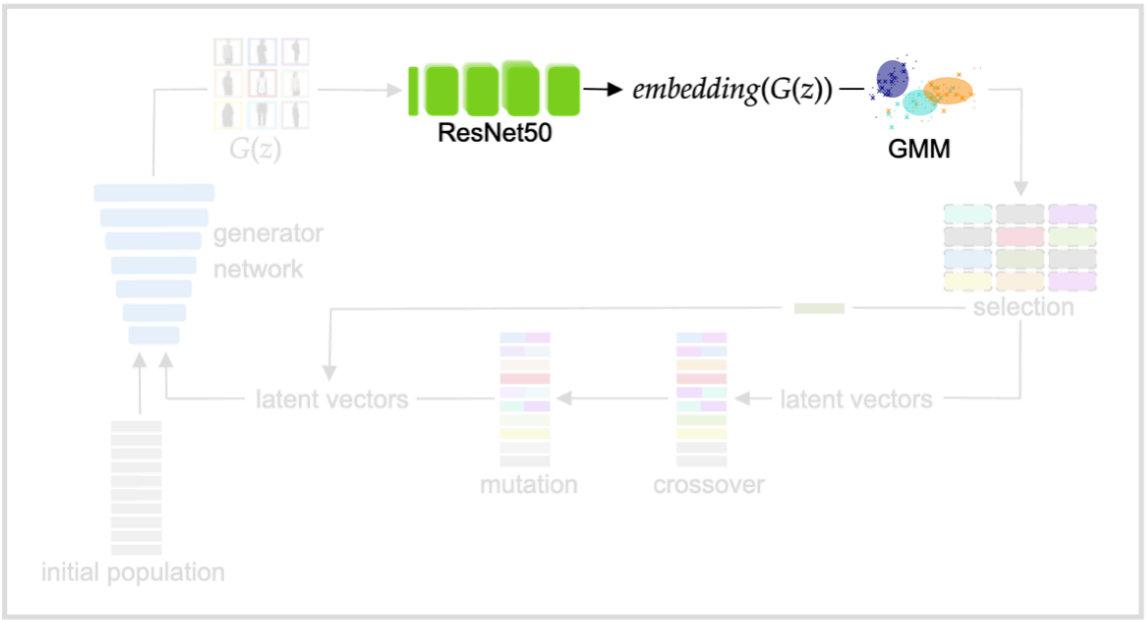


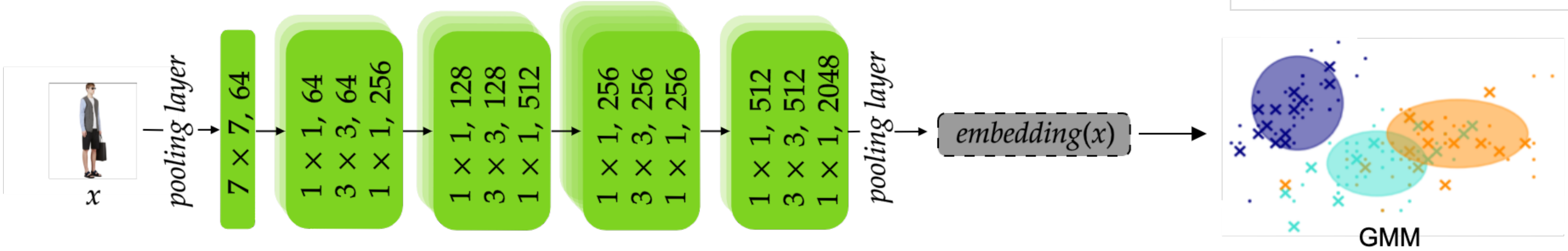


Residual Neural Network (ResNet)

- ResNet50 (Takagi et al., 2017)
- pre-trained on DeepFashion (Al-Halah and Grauman, 2020)
- penultimate embedding (2048-dim.) to extract high-level, but not class-specific “visual themes” (Matzen et al., 2017)

1) Style model





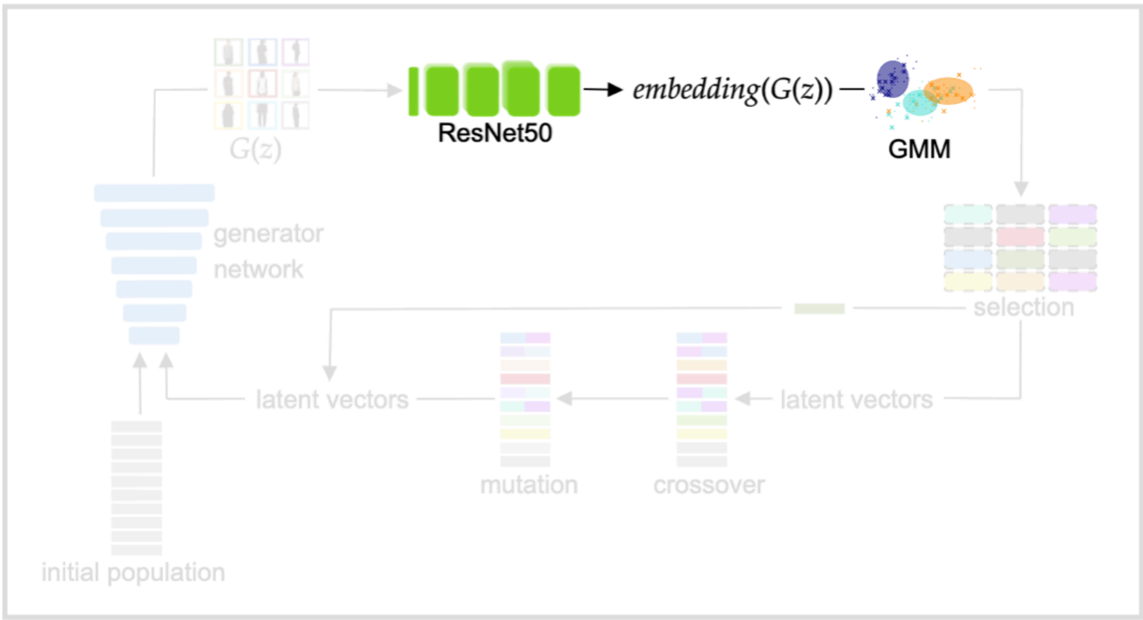
Residual Neural Network (ResNet)

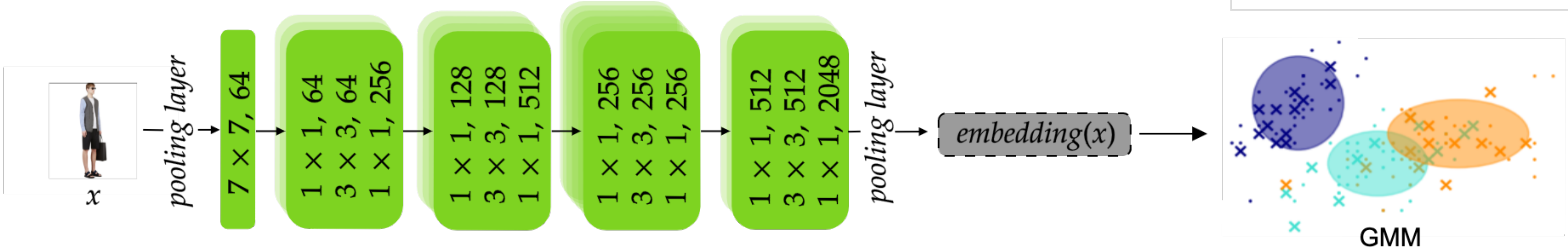
- ResNet50 (Takagi et al., 2017)
- pre-trained on DeepFashion (Al-Halah and Grauman, 2020)
- penultimate embedding (2048-dim.) to extract high-level, but not class-specific “visual themes” (Matzen et al., 2017)

Gaussian Mixture Model (GMM)

- diagonal covariance matrix to identify style components (Matzen et al., 2017)
- probabilistic model

1) Style model





Residual Neural Network (ResNet)

- ResNet50 (Takagi et al., 2017)
- pre-trained on DeepFashion (Al-Halah and Grauman, 2020)
- penultimate embedding (2048-dim.) to extract high-level, but not class-specific “visual themes” (Matzen et al., 2017)

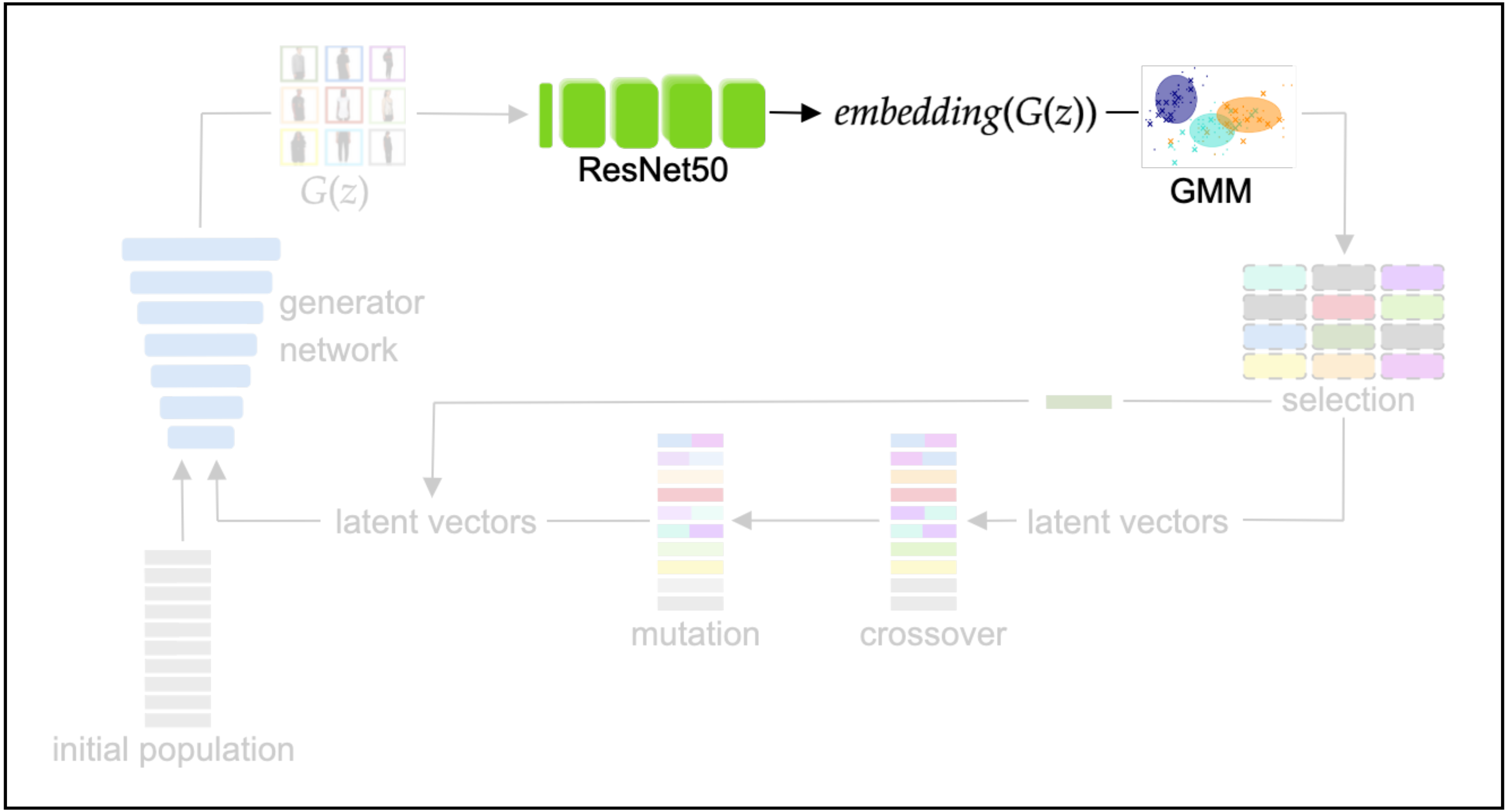
Gaussian Mixture Model (GMM)

- diagonal covariance matrix to identify style components (Matzen et al., 2017)
- probabilistic model
- posterior probability p_k of belonging to cluster component k

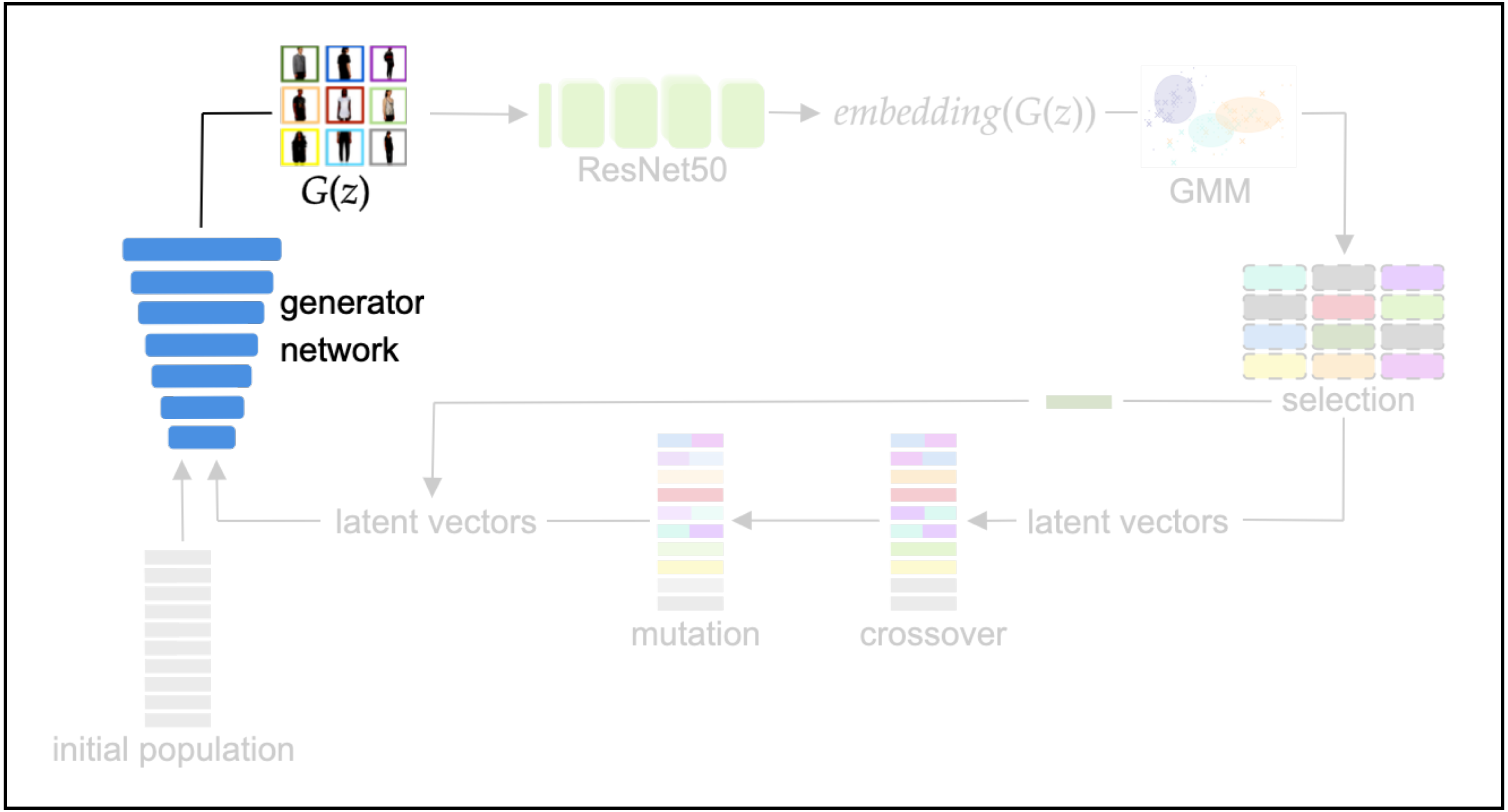
1) Style model

- Pose 4 images
- 150 style clusters

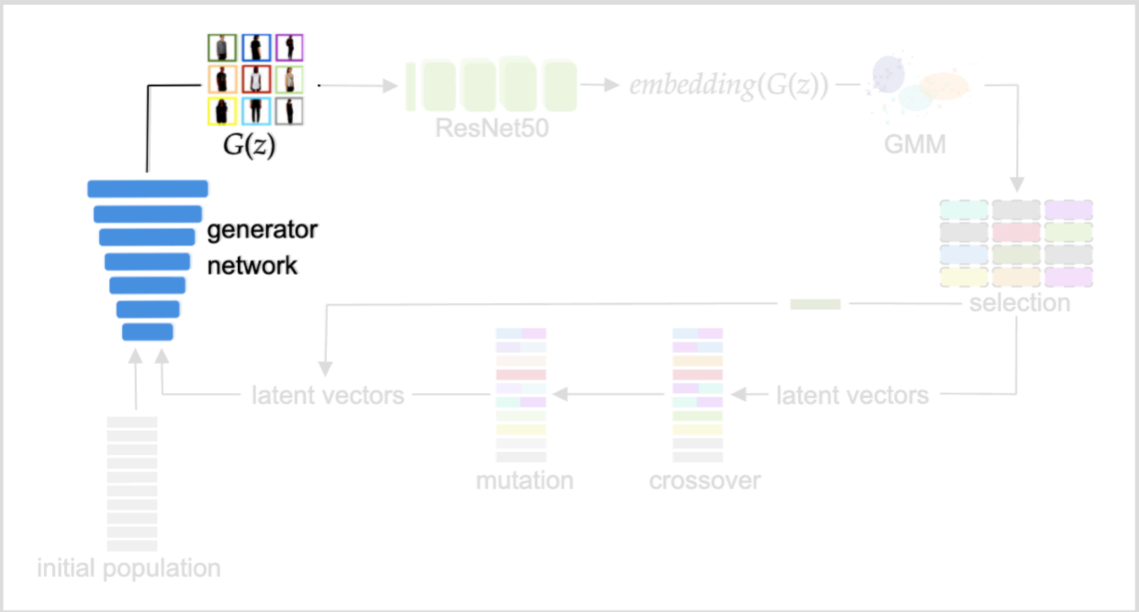
1) Style model

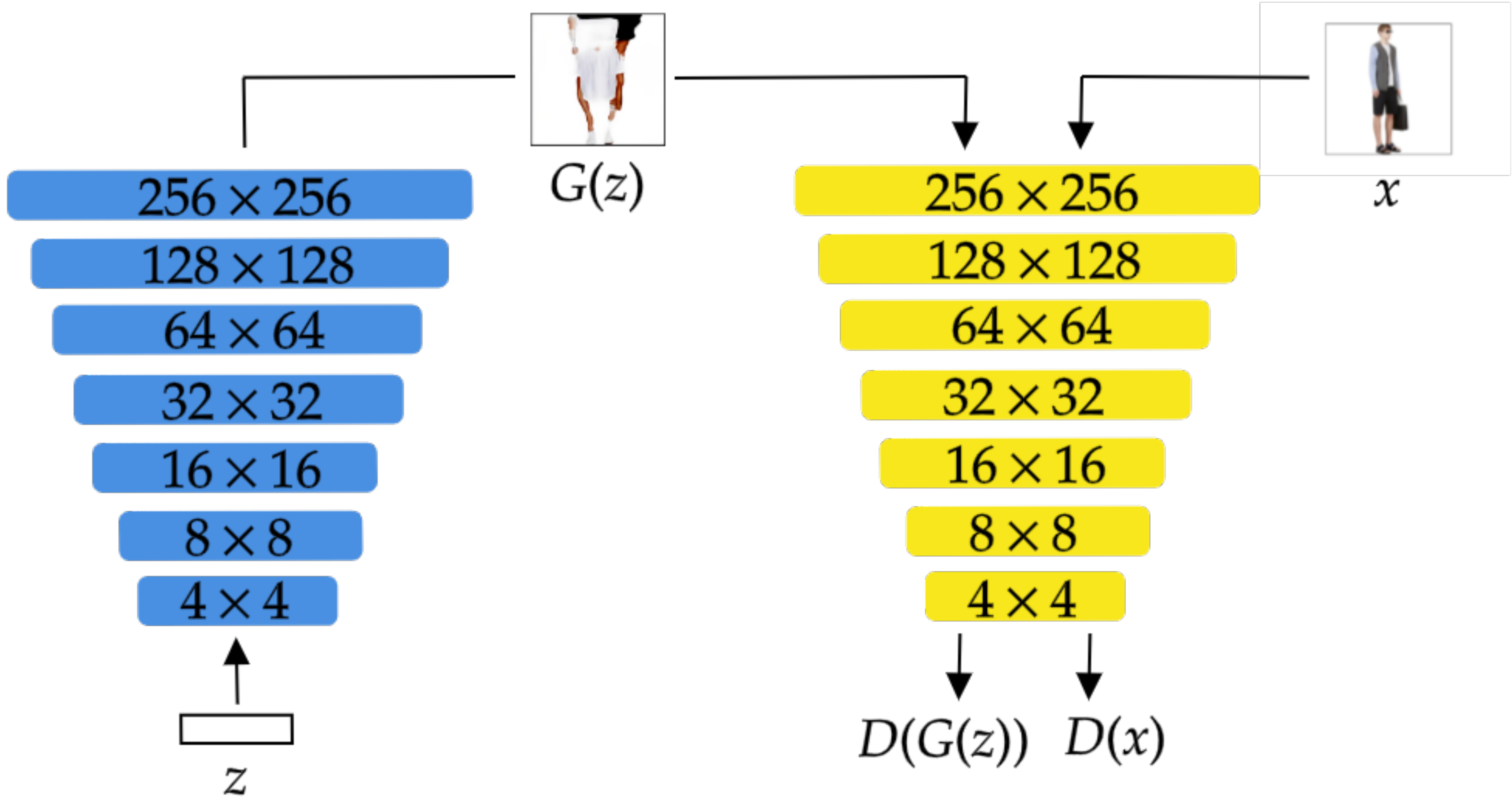


2) Generative model



2) Generative model

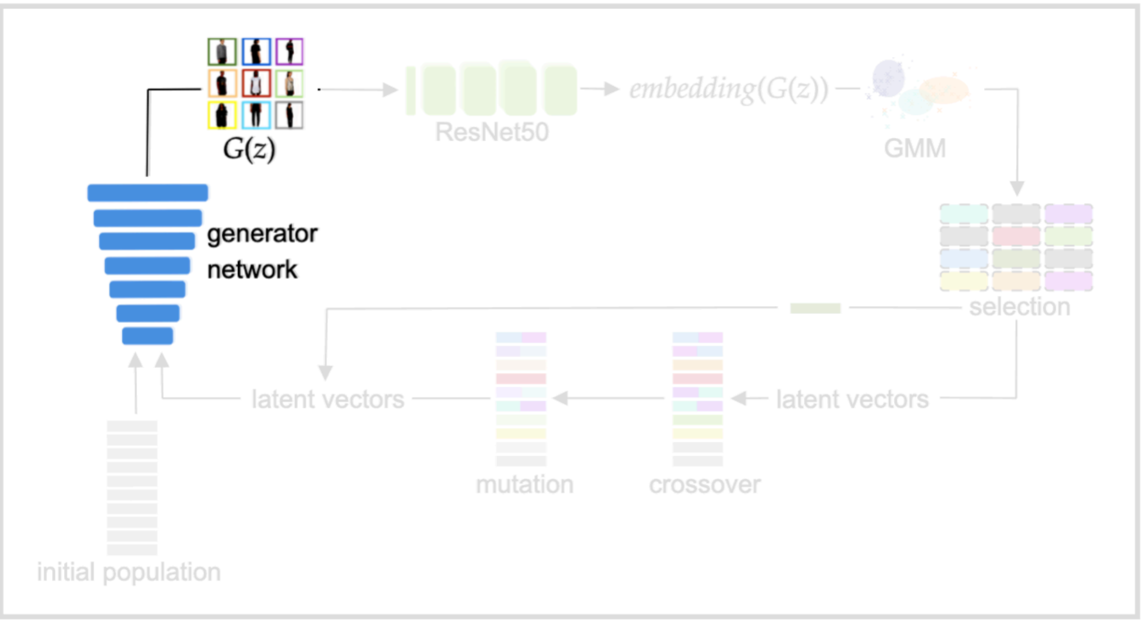


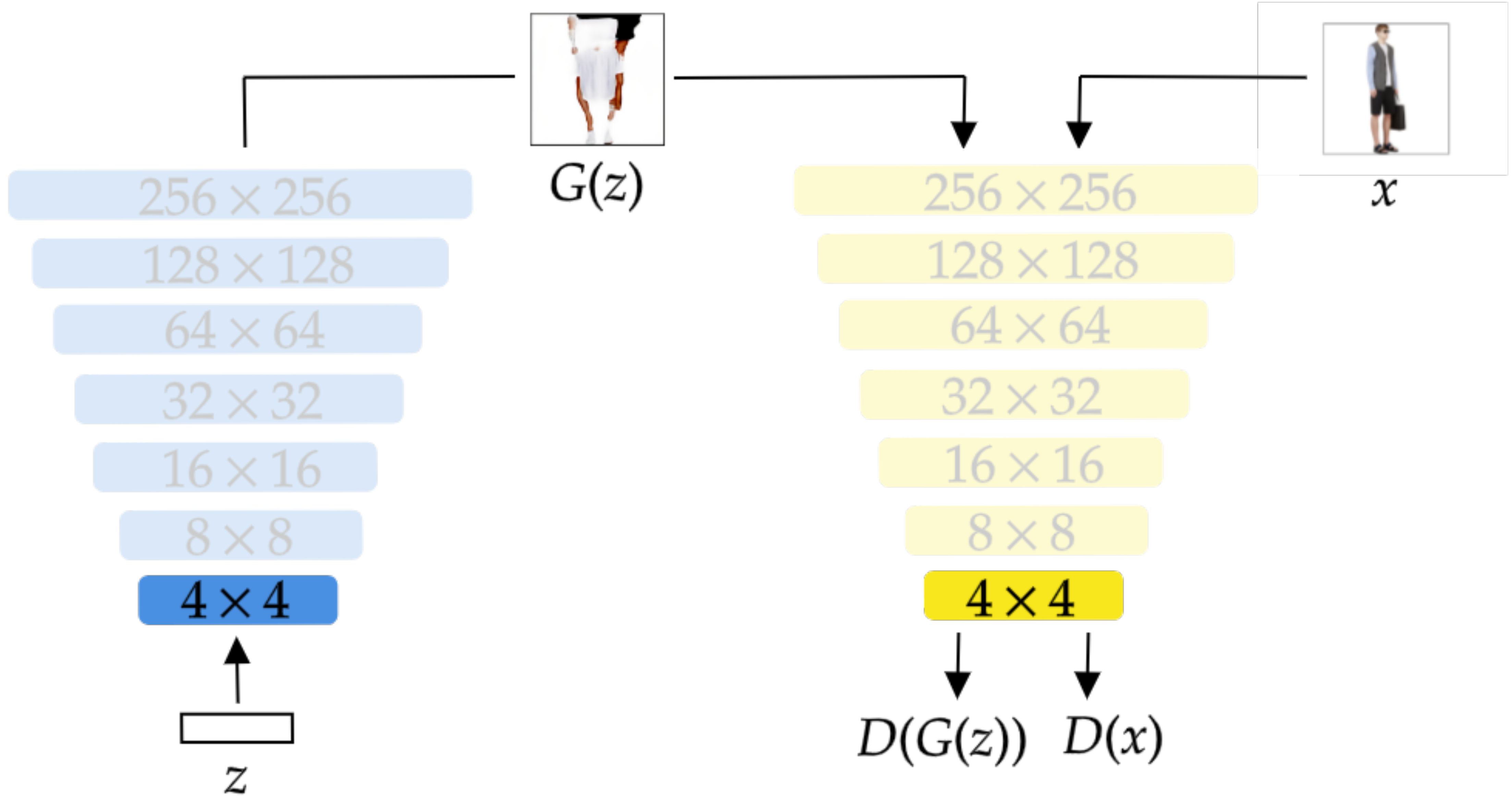


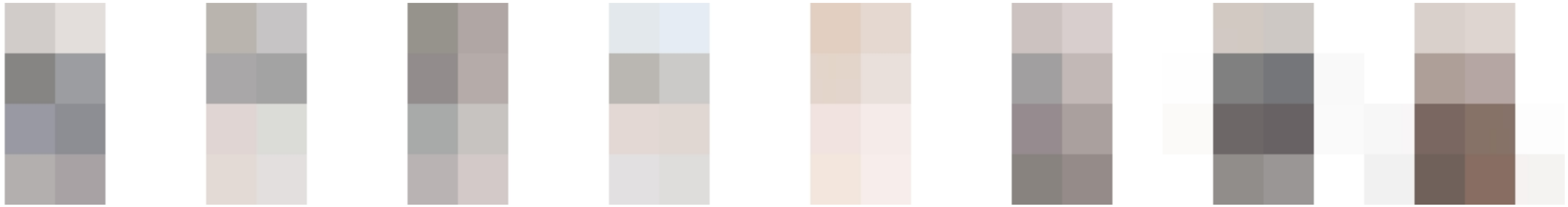
Progressively growing GANs (P-GANs) (Karras et al., 2017)

- latent vector $z \sim \mathcal{N}(\mu = 0, \sigma = 1)$ of length 512
- Wasserstein GAN loss with gradient penalty (WGAN-GP)

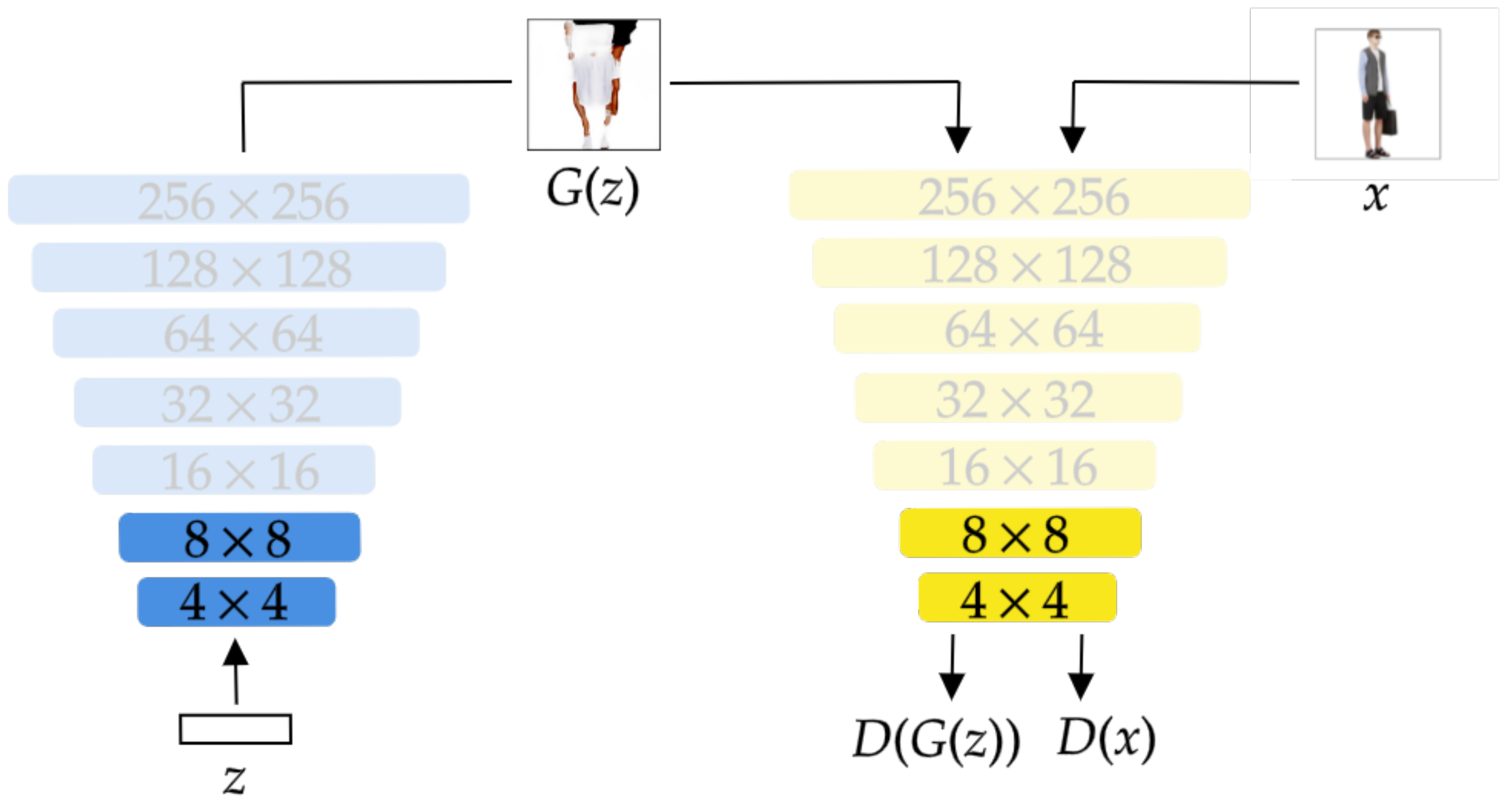
2) Generative model

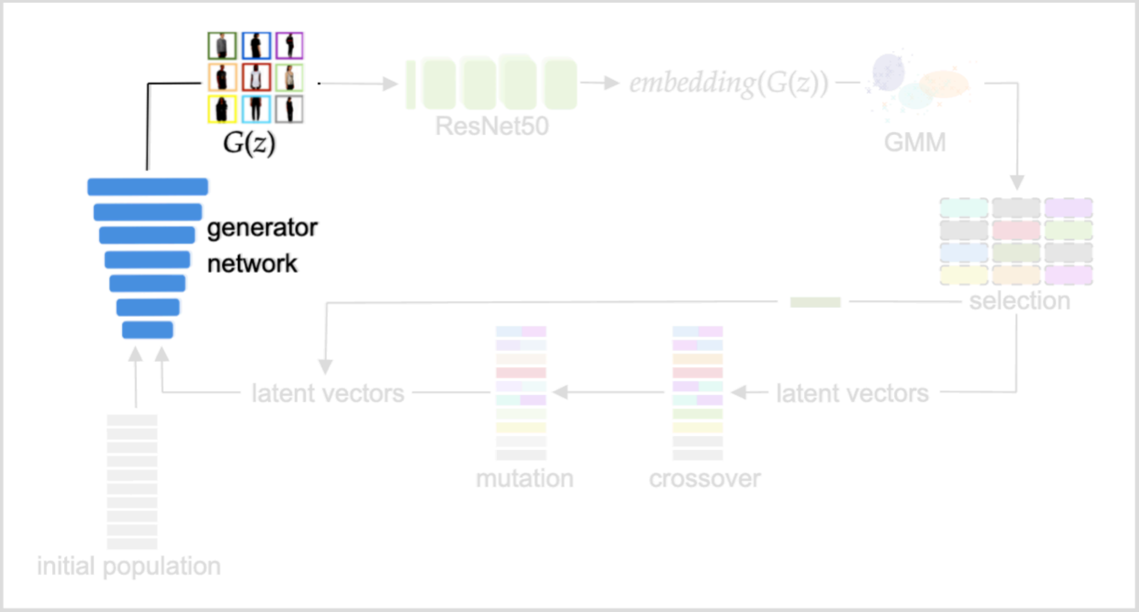


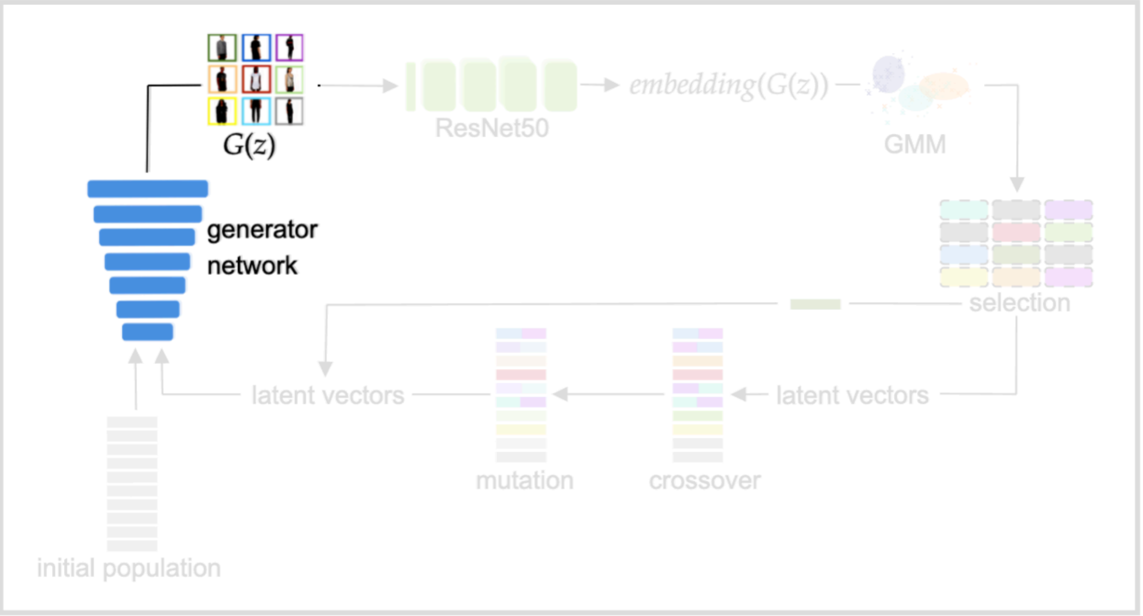




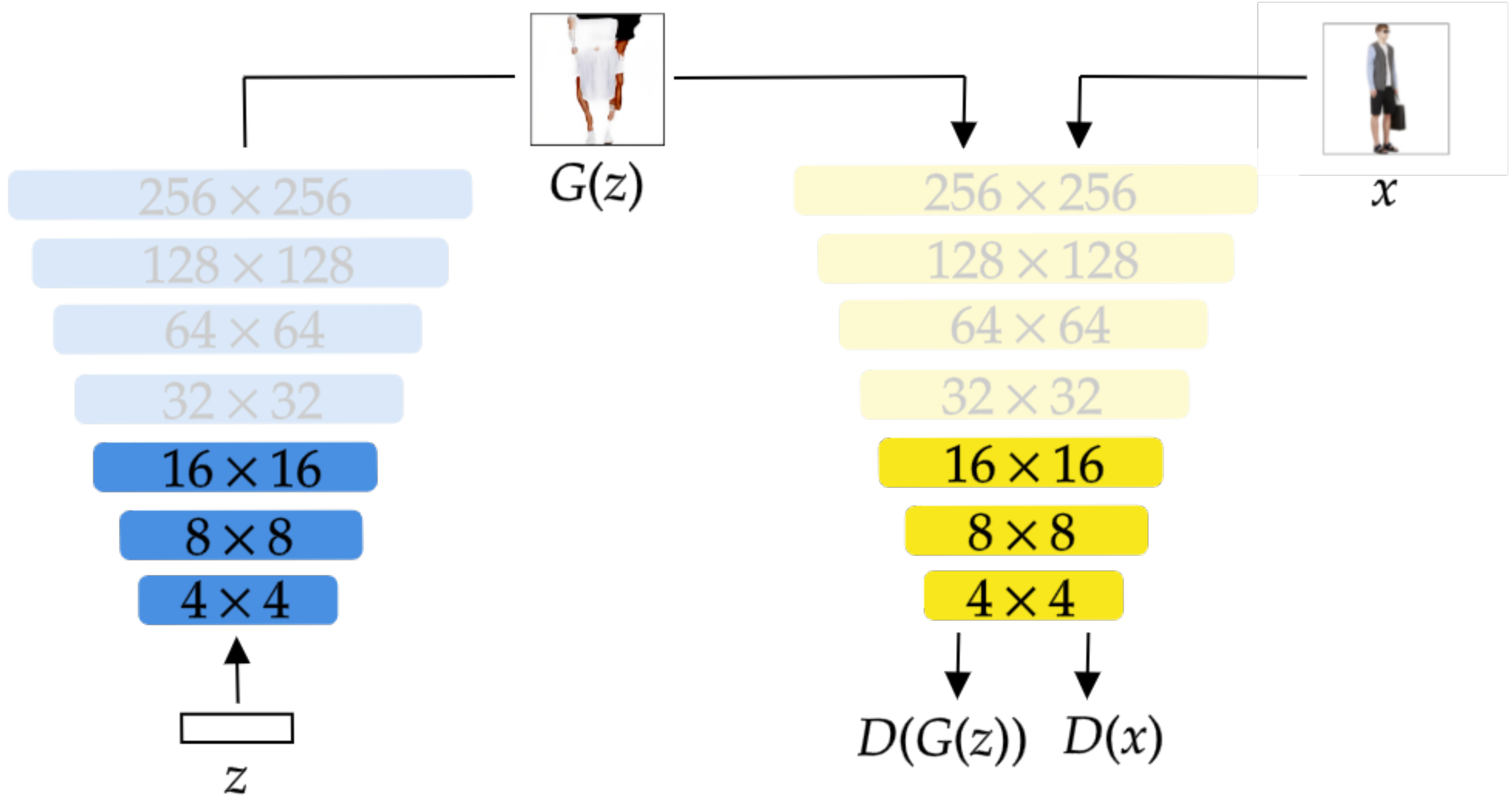
2) Generative model



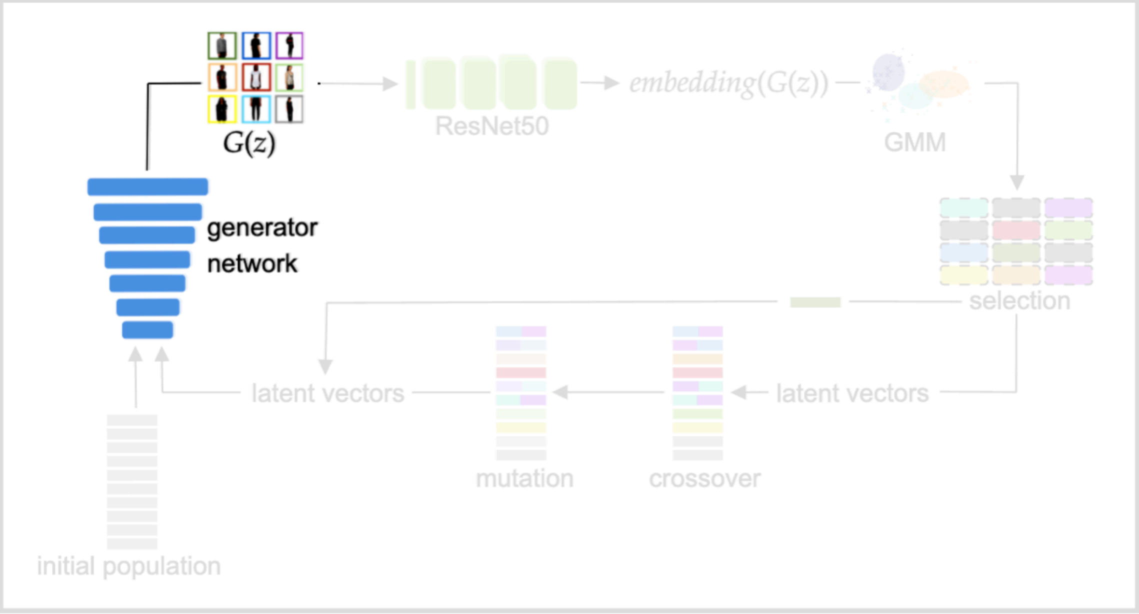


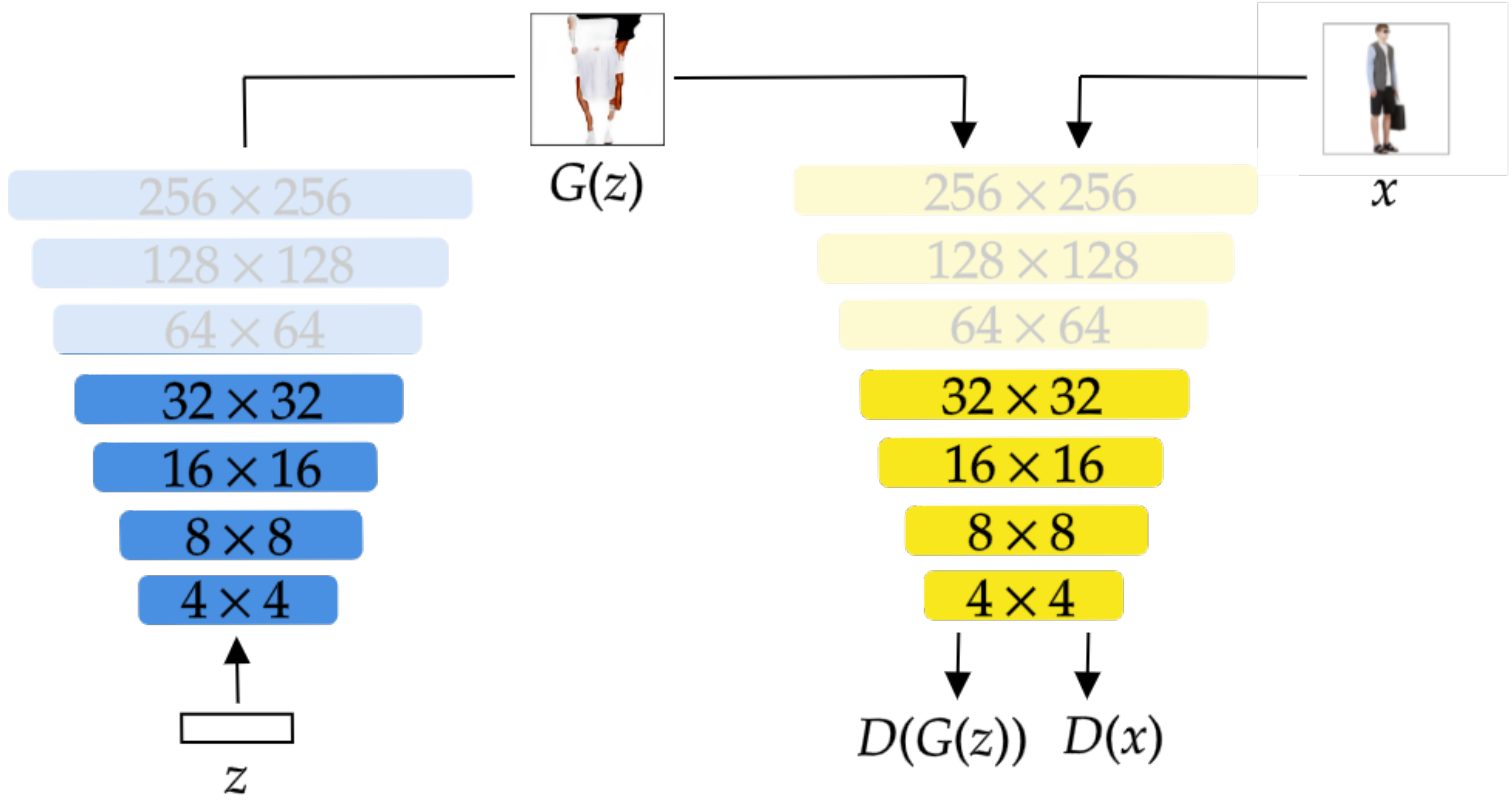


2) Generative model

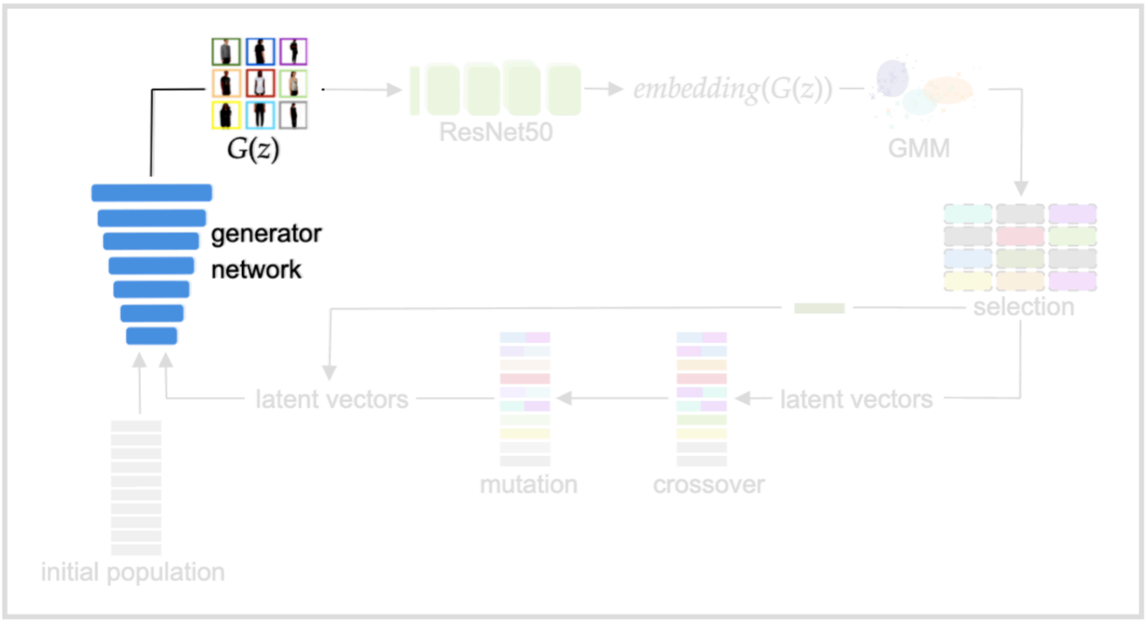


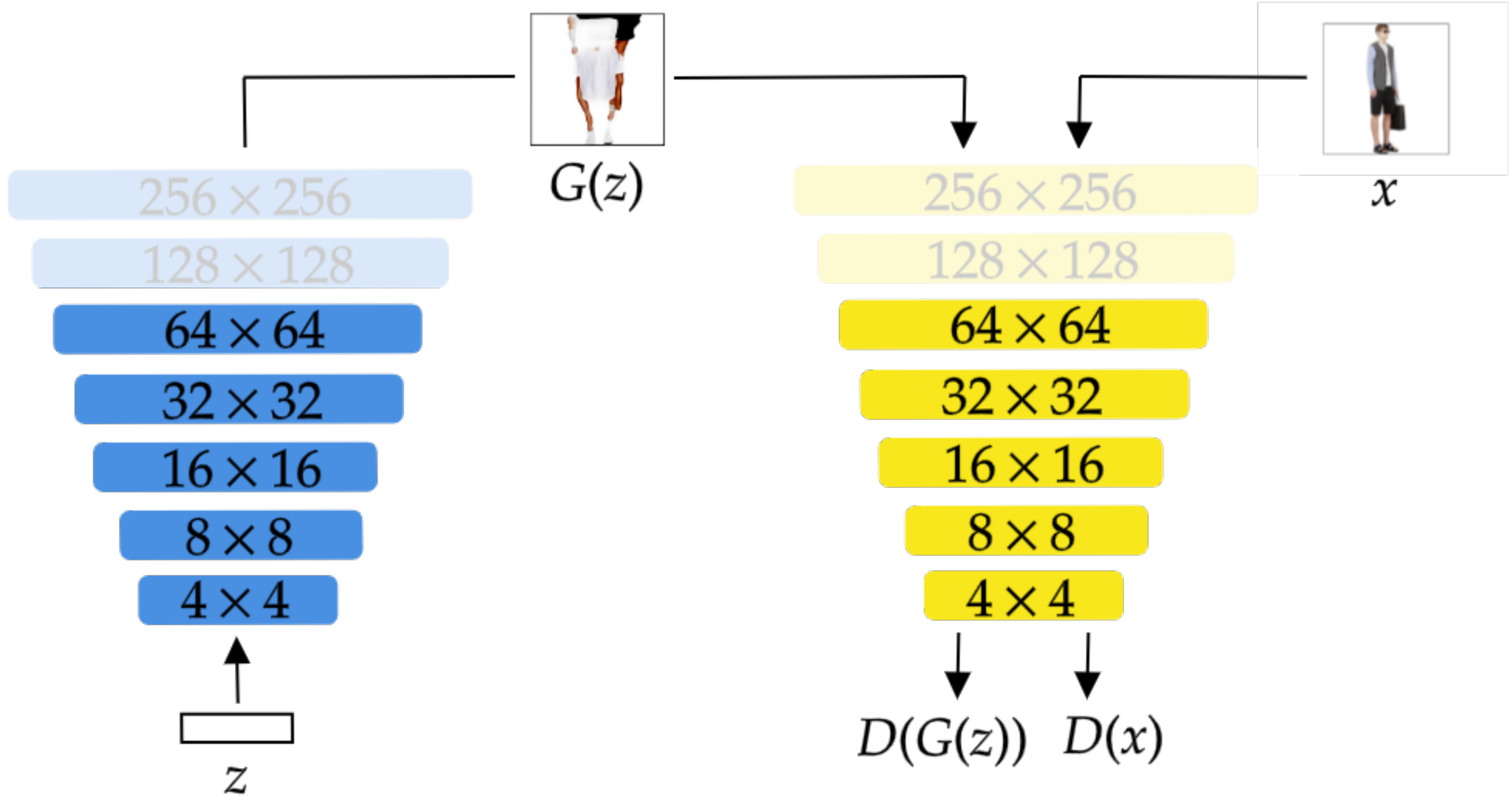
2) Generative model



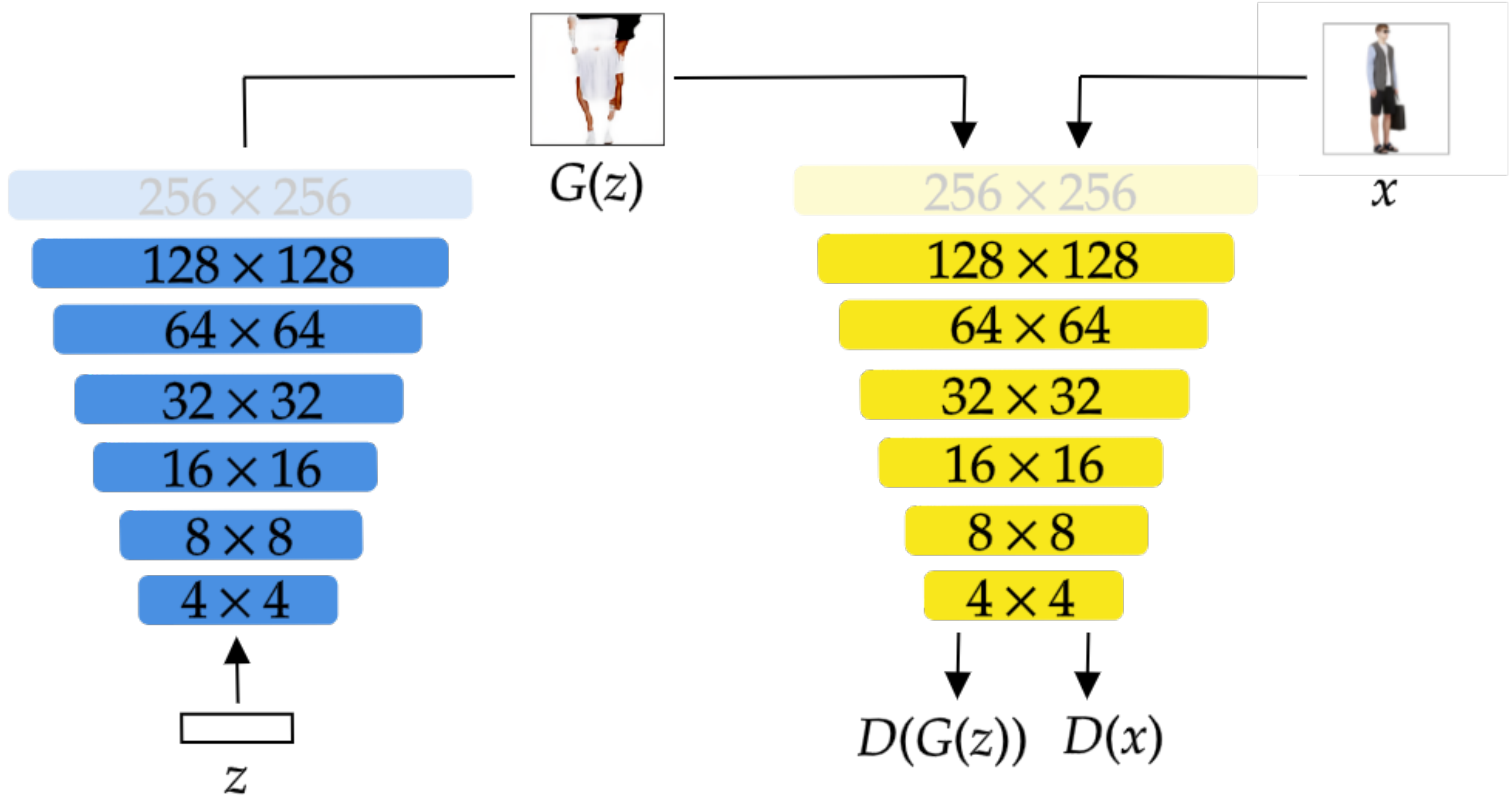


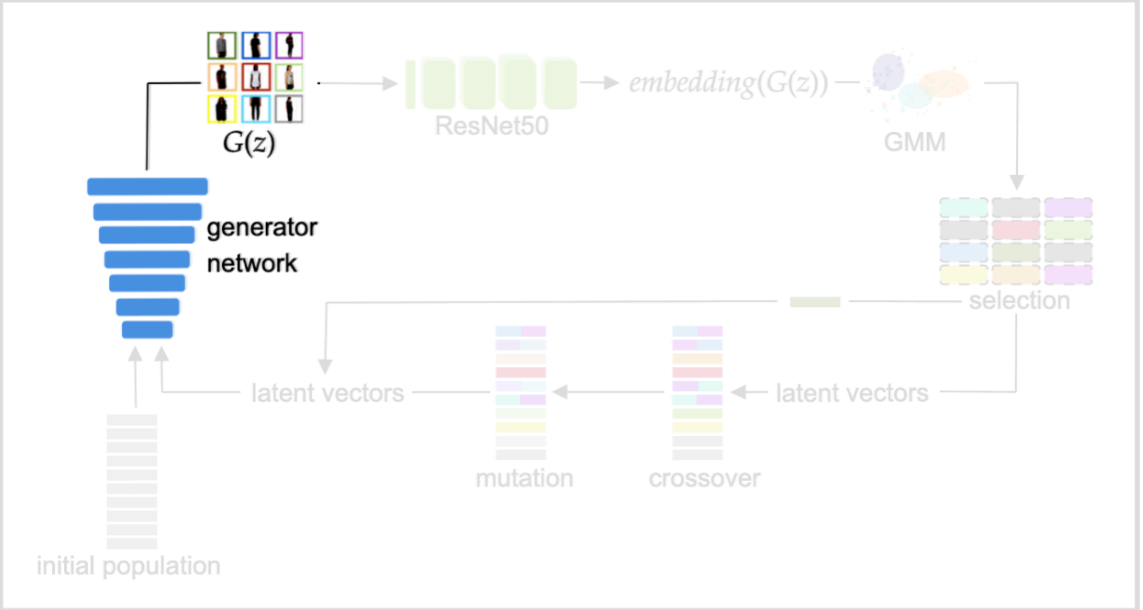
2) Generative model



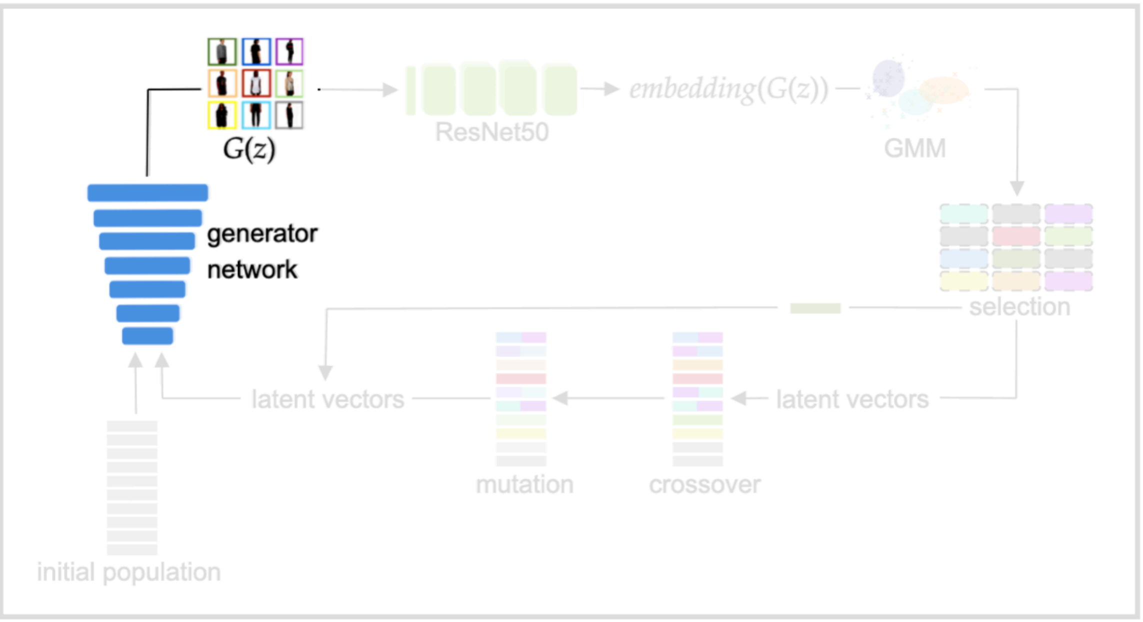


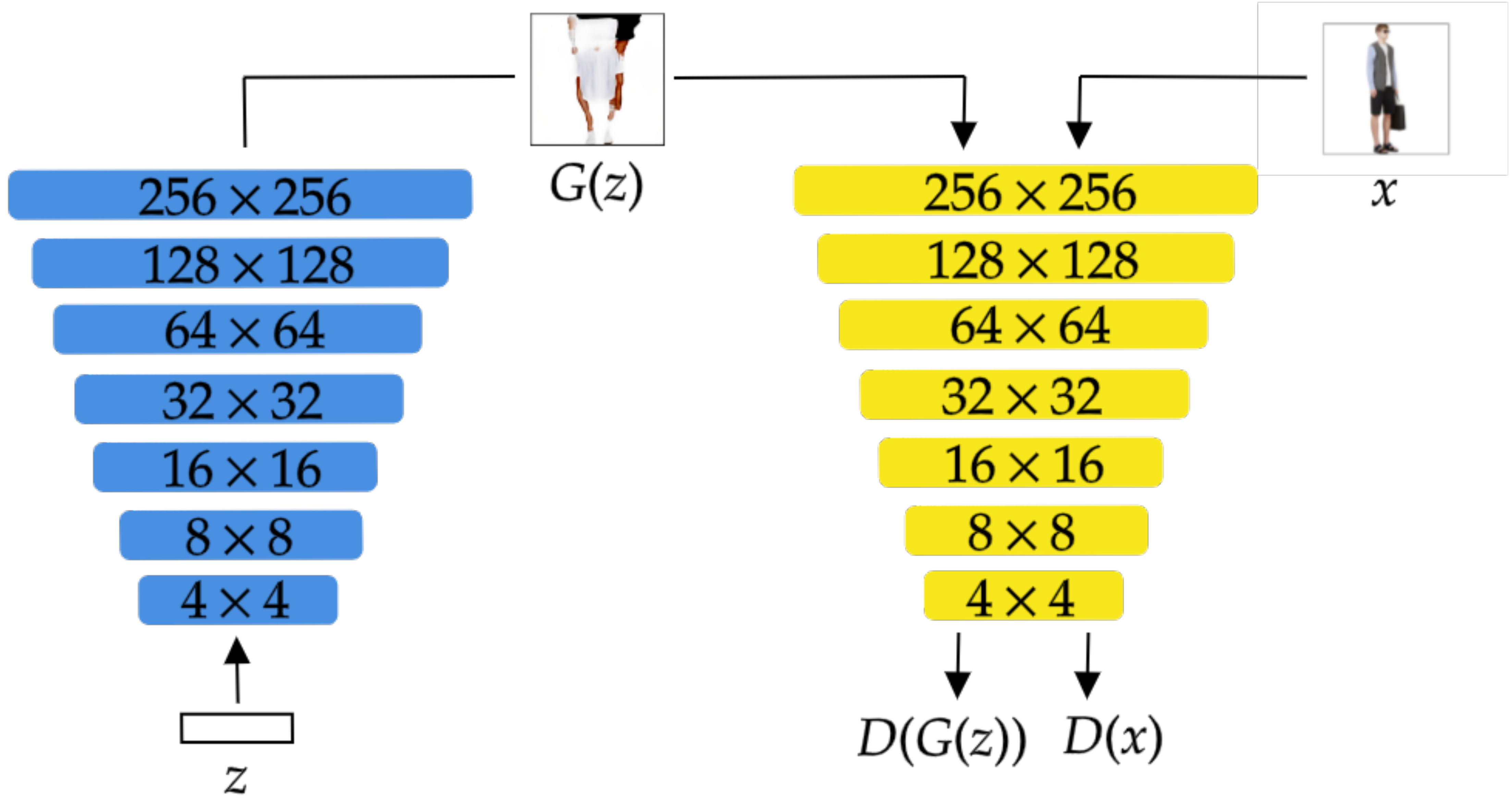
2) Generative model



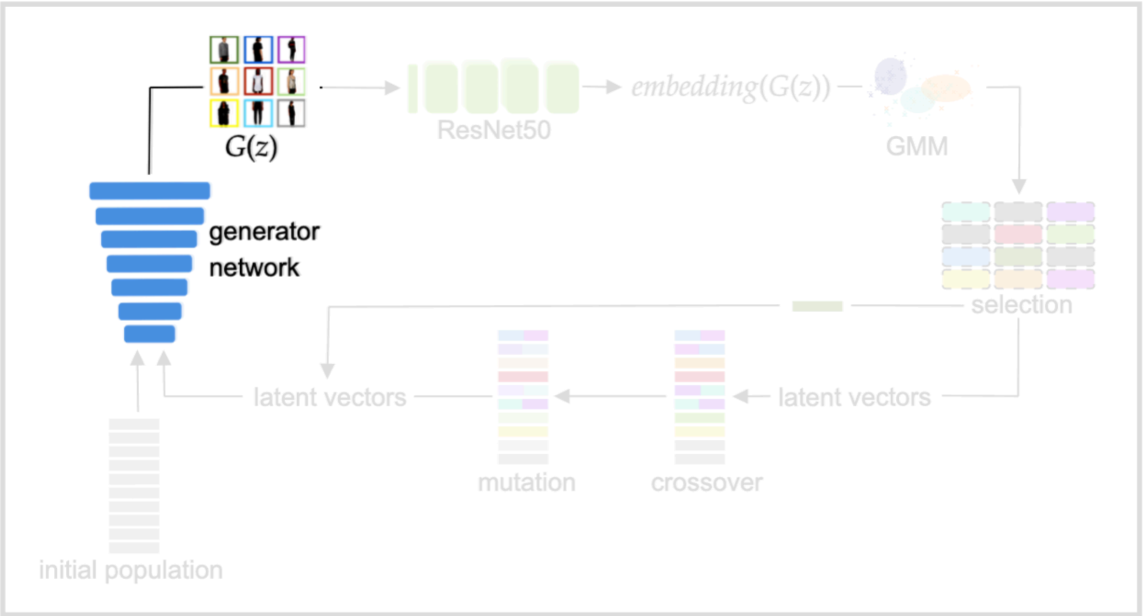


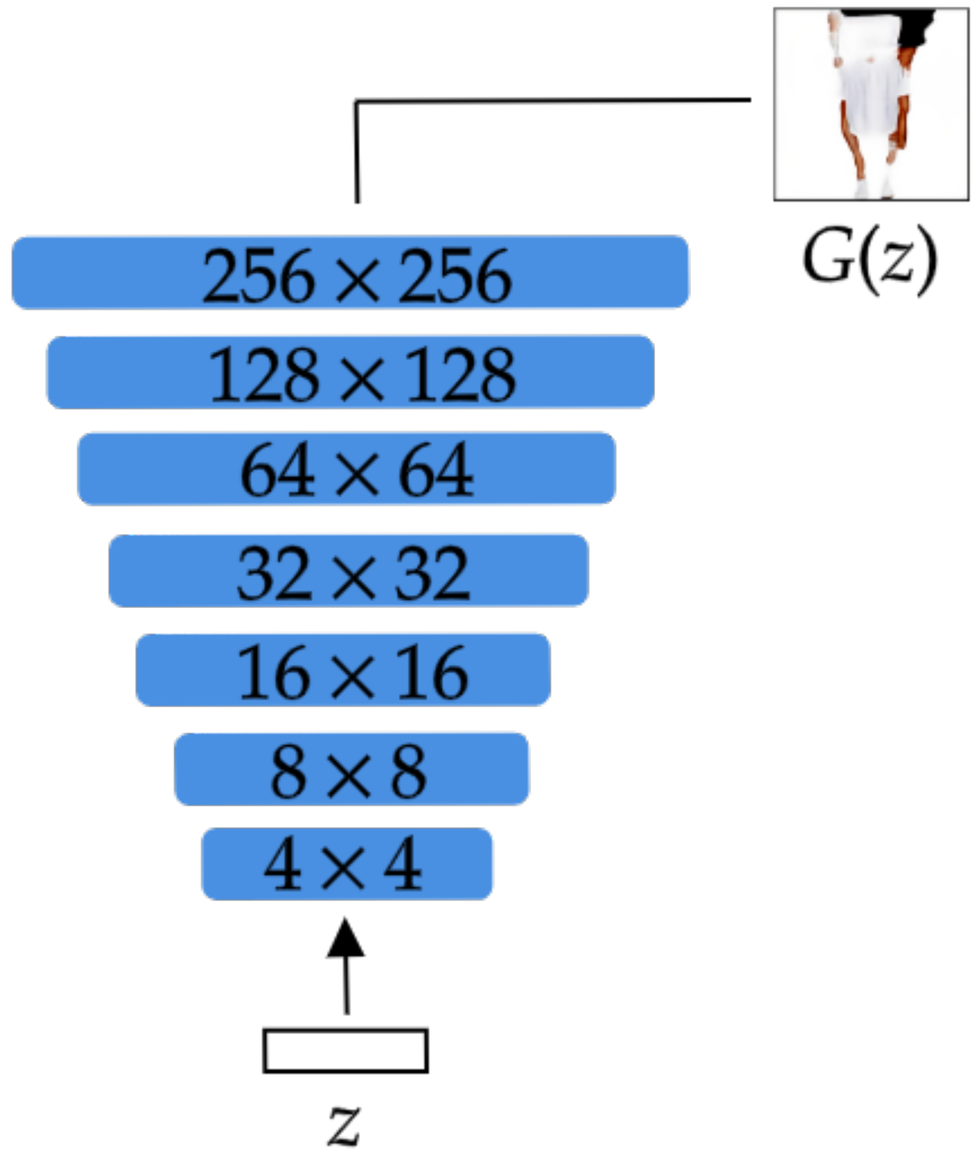
2) Generative model



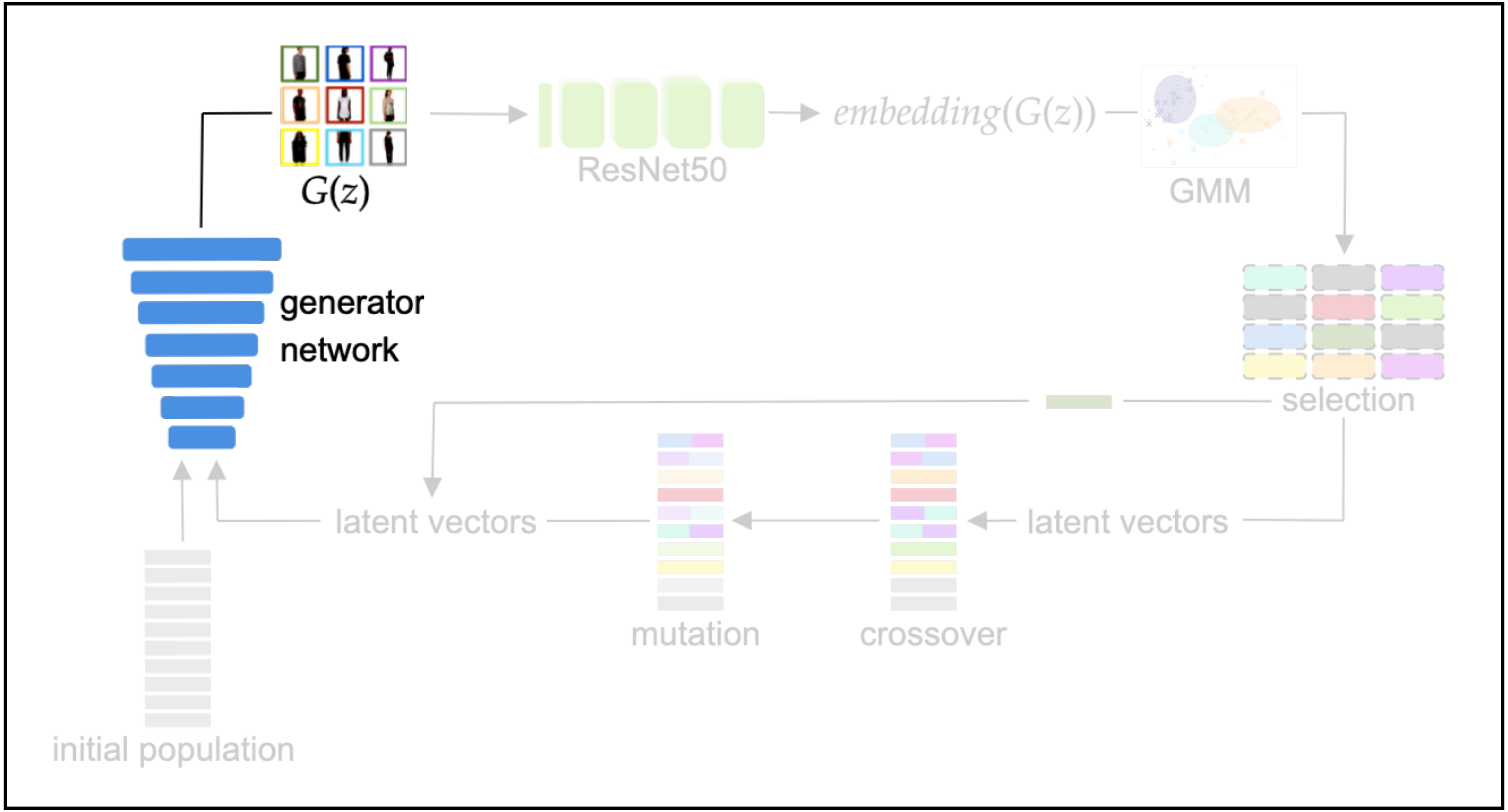


2) Generative model

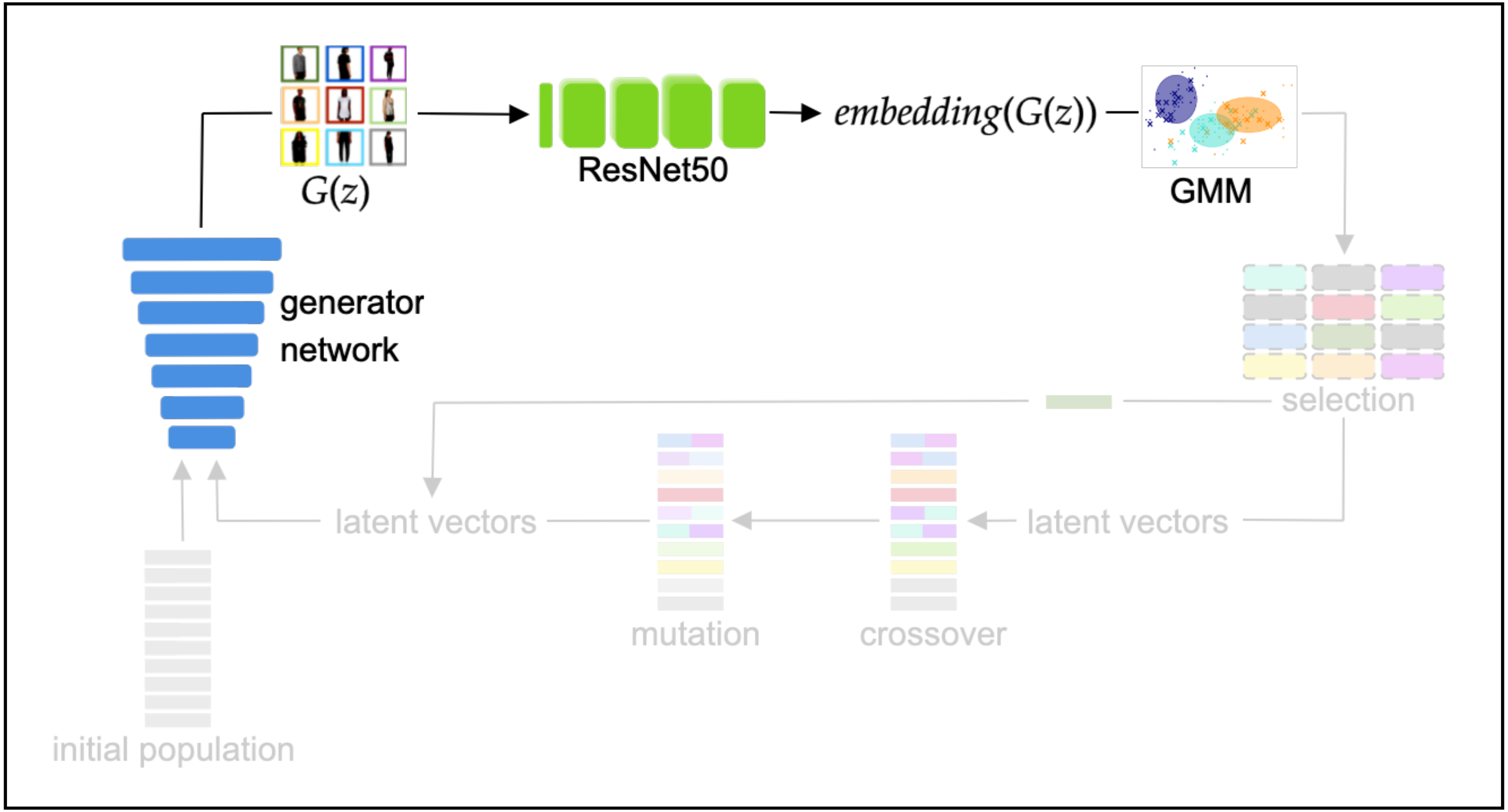




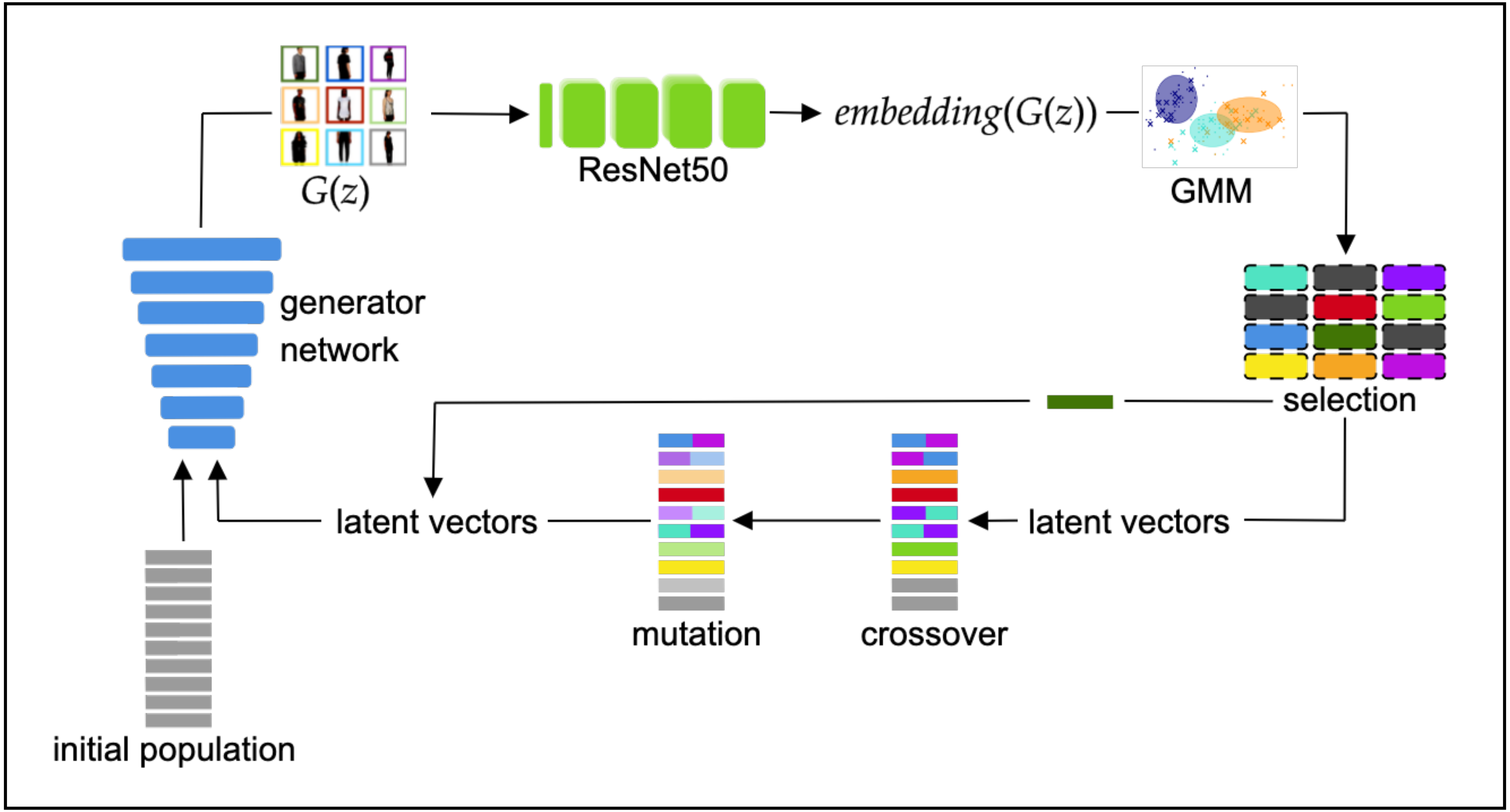
2) Generative model



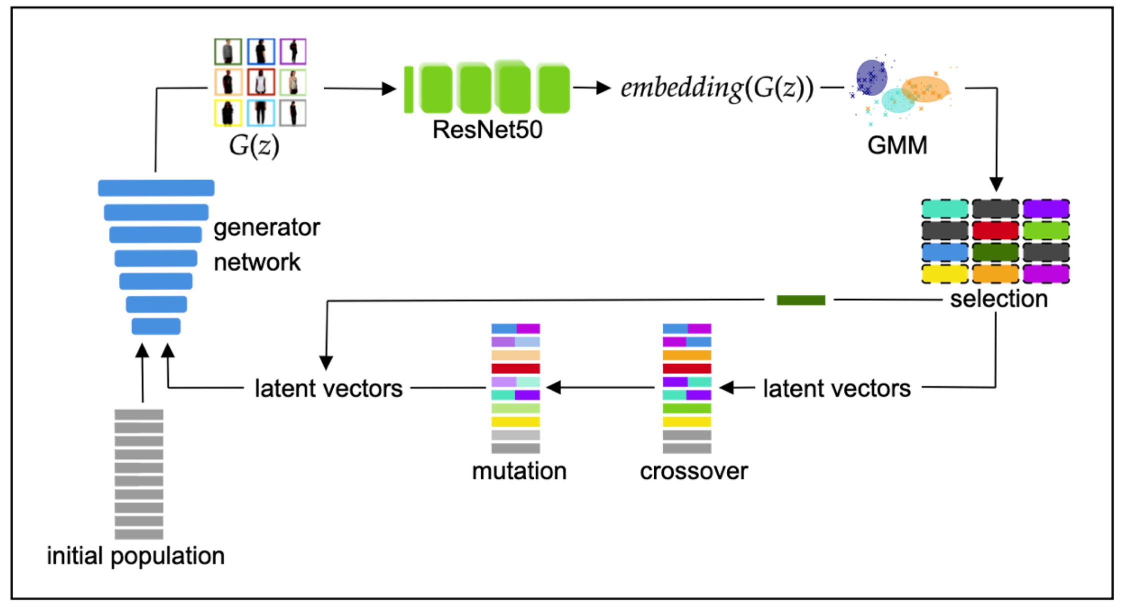
2) Generative model



3) Genetic algorithm



3) Genetic algorithm (GA)

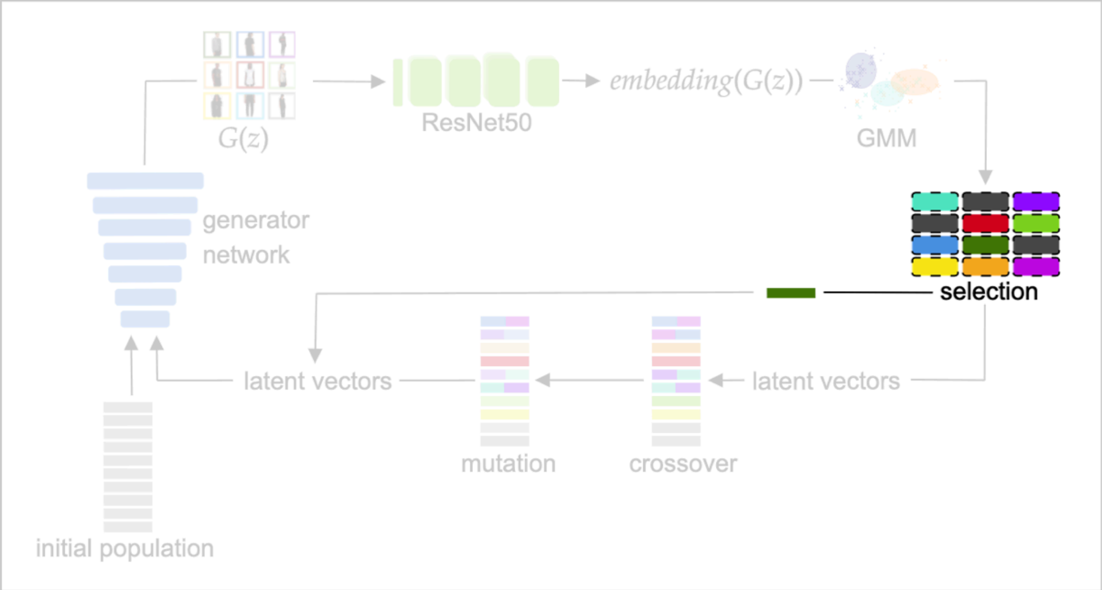


Representation:

- latent vector $z = \langle v_1, \ldots, v_l \rangle$ with $v \in \mathbb{R}$ is an individual
- population of N_{pop} individuals

Fitness F_t with regard to target style t: $f_t(z) = p_t(embedding(G(z)))$

Objective: $z* = \arg\max_z \; F_t(z)$



3) Genetic algorithm (GA)

Representation:

- latent vector $z = \langle v_1, \ldots, v_l \rangle$ with $v \in \mathbb{R}$ is an individual
- population of N_{pop} individuals

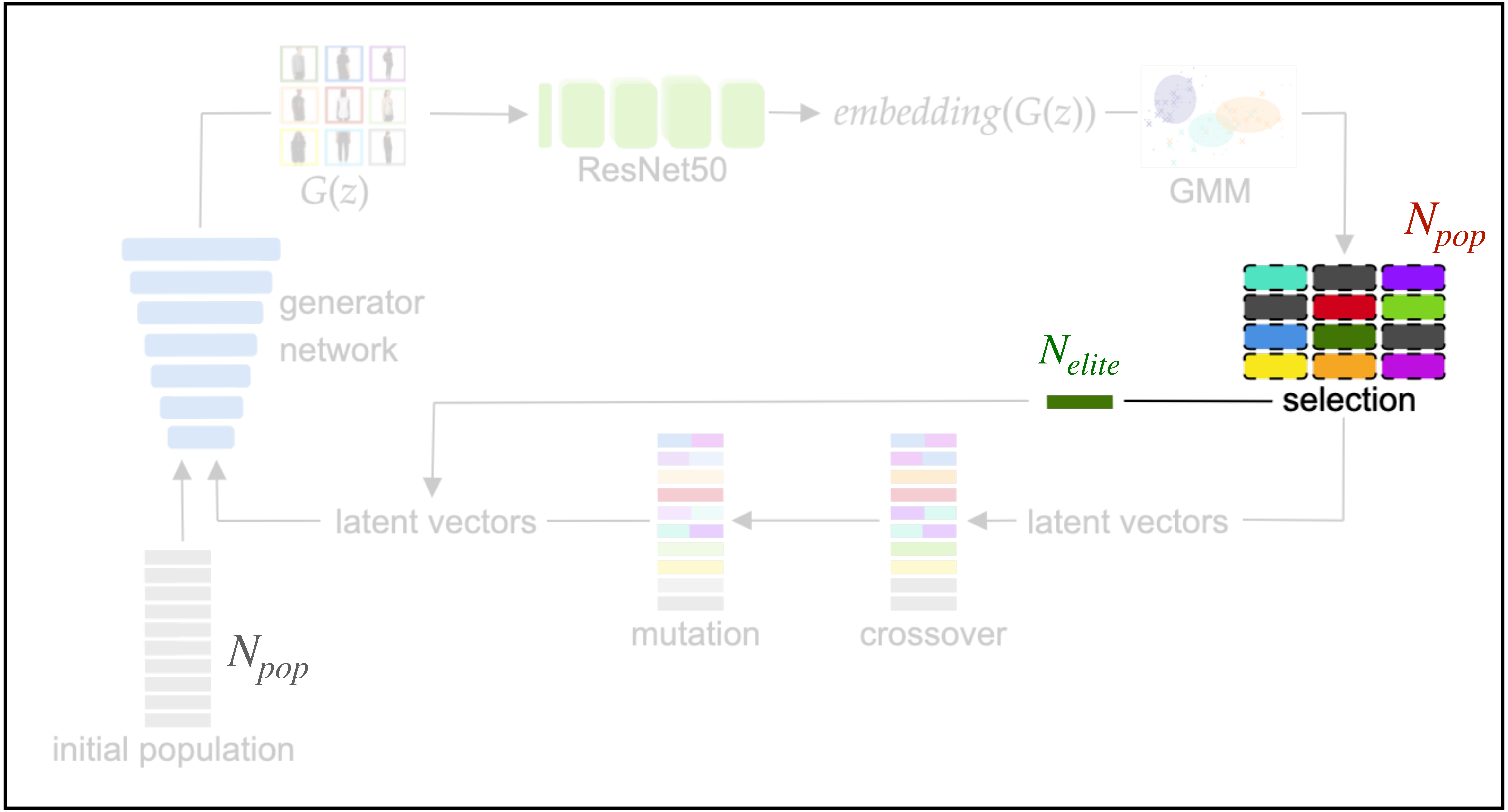
Fitness F_t with regard to target style t: $f_t(z) = p_t(embedding(G(z)))$

Objective: $z^* = \arg\max_z \; F_t(z)$

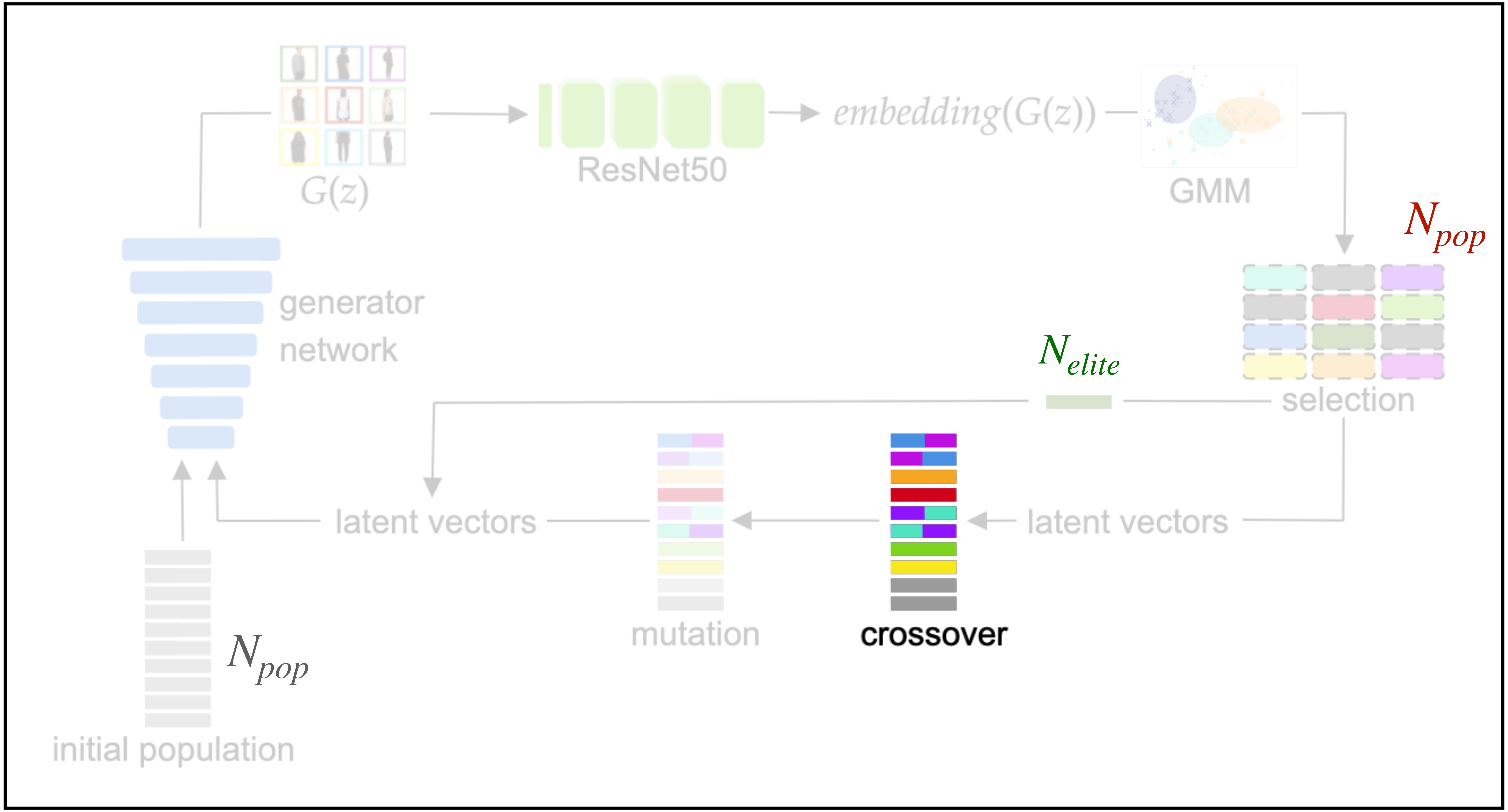
Selection:

- Tournaments of size N_{ts} to choose N_{pop} individuals for the next generation
- additional N_{elite} best individual(s) preserved for next generation

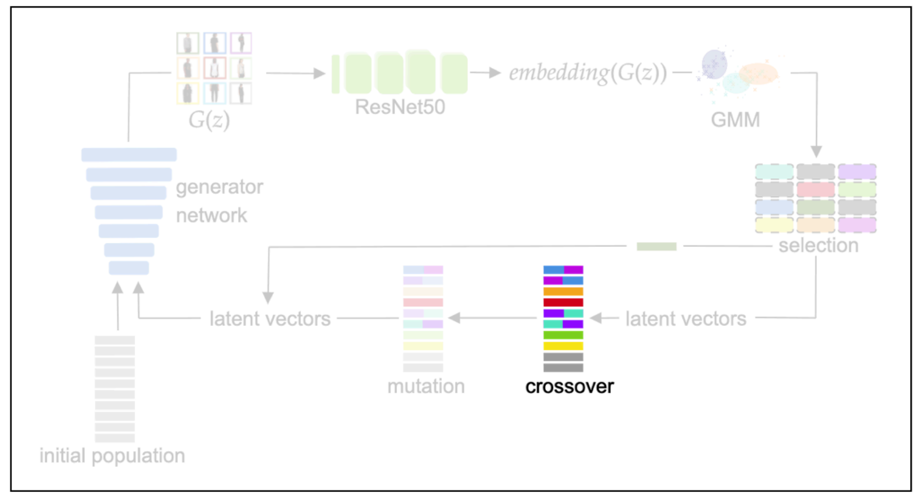
3) Genetic algorithm



3) Genetic algorithm



3) Genetic algorithm



Uniform crossover between replaces two individuals a and b (parents) with the children:

$$\hat{a}_i = \alpha a_i + (1-\alpha) b_i \quad \text{and} \quad \hat{b}_i = \alpha b_i + (1-\alpha) a_i \qquad \text{with } \alpha = Bernoulli(0.5)$$

0.5	-0.7	0.1	0.3	1.1	-1	-0.2	0.4	-2	0.6

0.7	1.2	1.5	0	-0.2	0.5	-0.9	-1.1	-0.1	2.3

↓

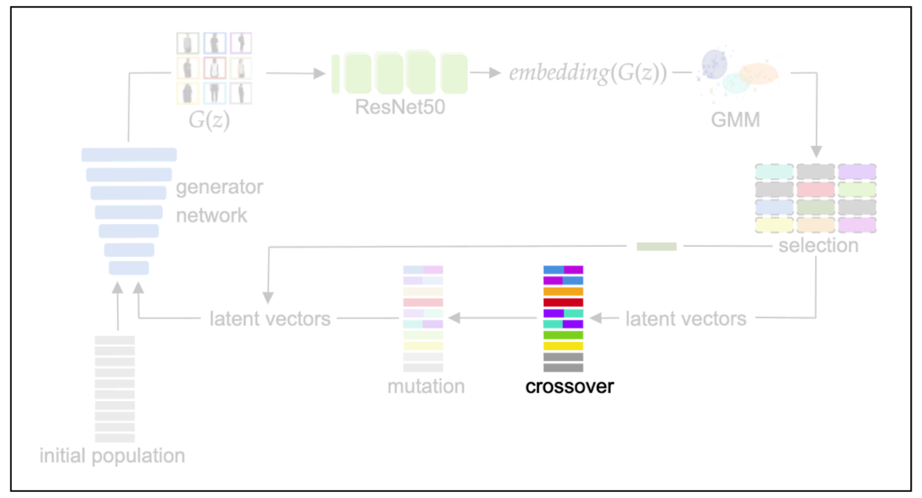
0.5	1.2	1.5	0.3	1.1	0.5	-0.2	-1.1	-2	2.3

0.7	1.2	0.1	0	1.1	0.5	-0.2	0.4	-2	2.3

An individual participates in crossover by chance of crossover rate p_{cx}.

N_{new} individuals added to introduce new gene material.

3) Genetic algorithm



Uniform crossover between replaces two individuals a and b (parents) with the children:

$$\hat{a}_i = \alpha a_i + (1-\alpha) b_i \quad \text{and} \quad \hat{b}_i = \alpha b_i + (1-\alpha) a_i \qquad \text{with } \alpha = Bernoulli(0.5)$$

0.5	-0.7	0.1	0.3	1.1	-1	-0.2	0.4	-2	0.6

0.7	1.2	1.5	0	-0.2	0.5	-0.9	-1.1	-0.1	2.3

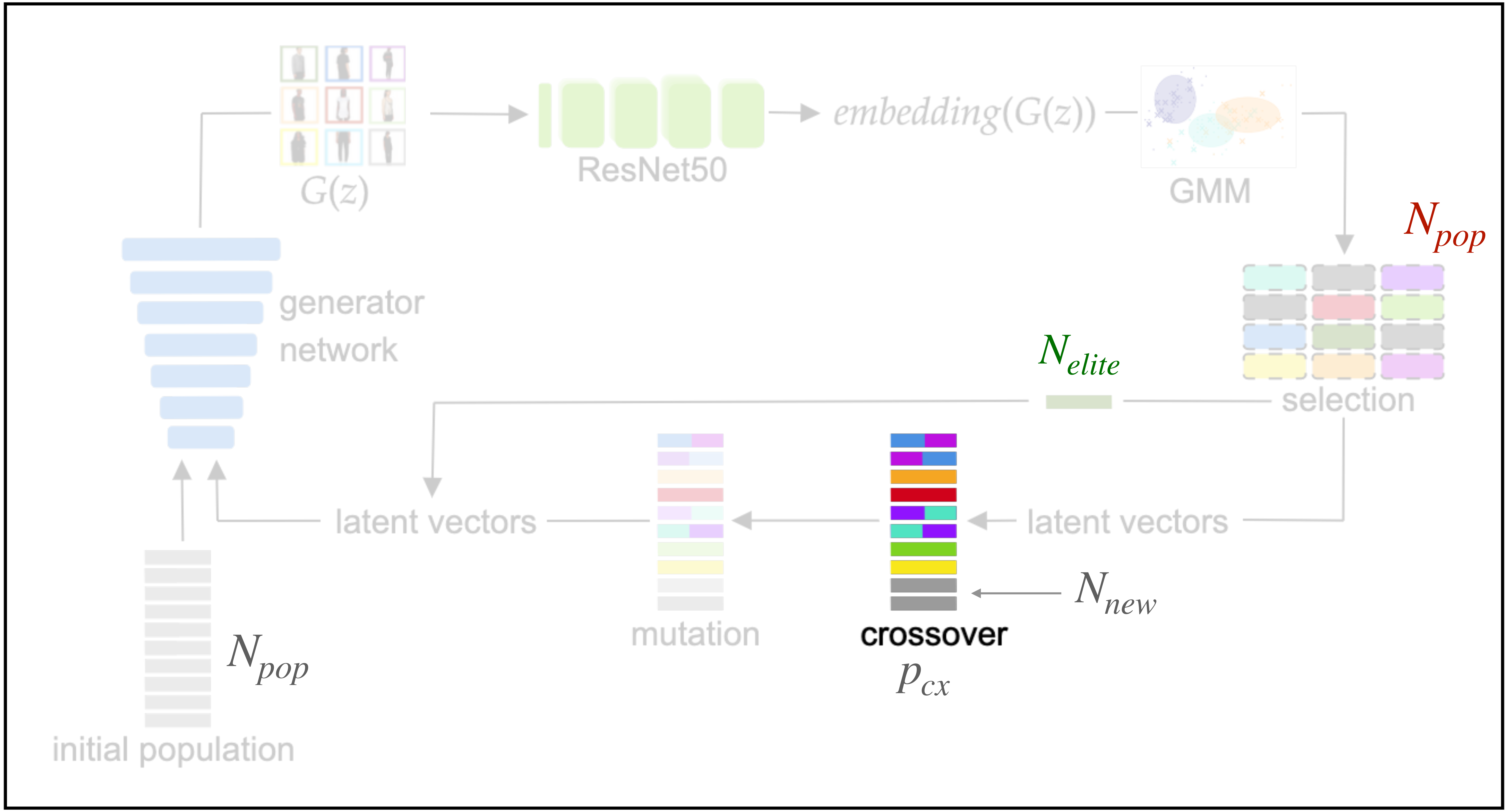
0.5	1.2	1.5	0.3	1.1	0.5	-0.2	-1.1	-2	2.3

0.7	1.2	0.1	0	1.1	0.5	-0.2	0.4	-2	2.3

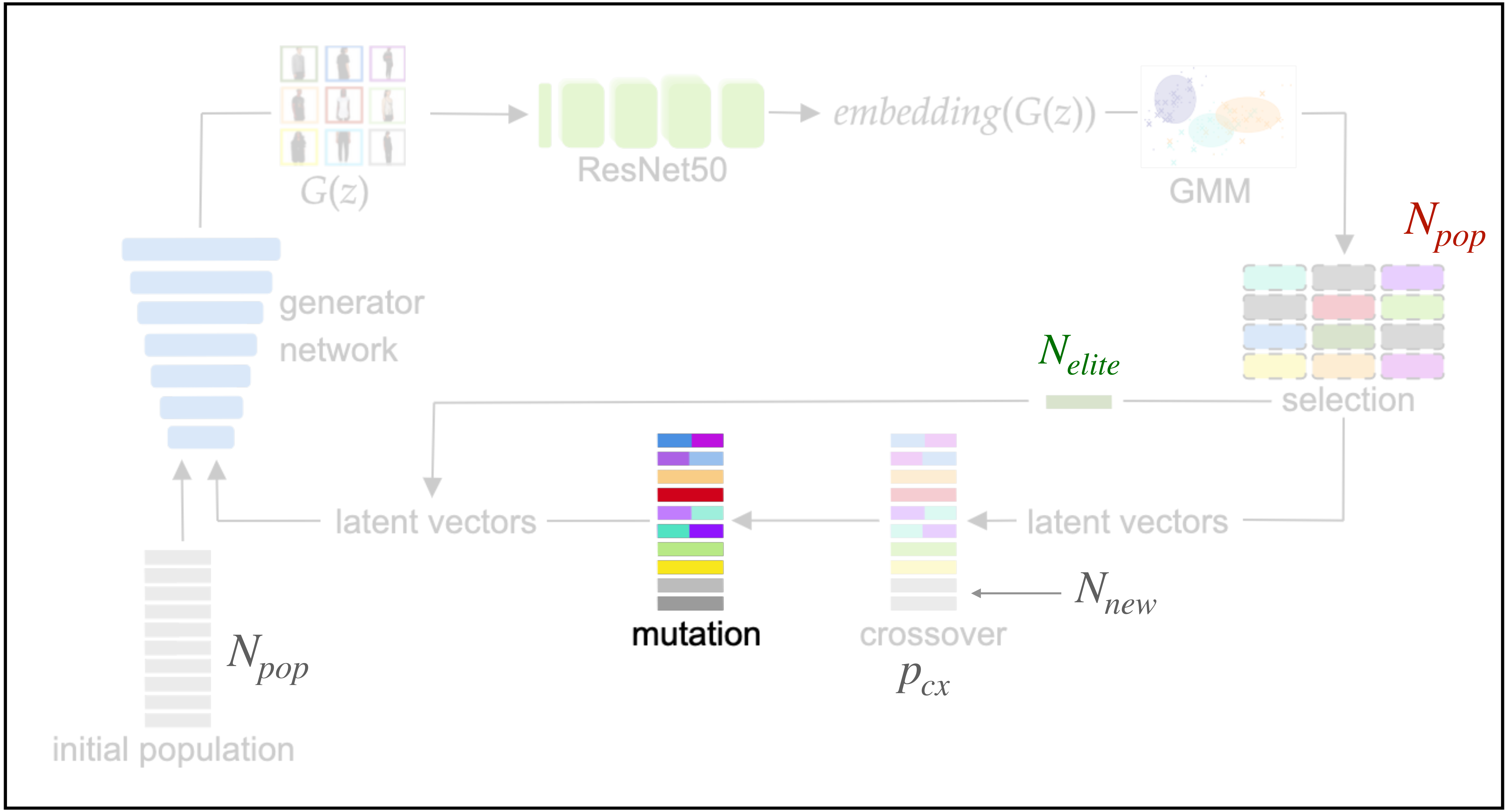
An individual participates in crossover by chance of crossover rate p_{cx}.

N_{new} individuals added to introduce new gene material.

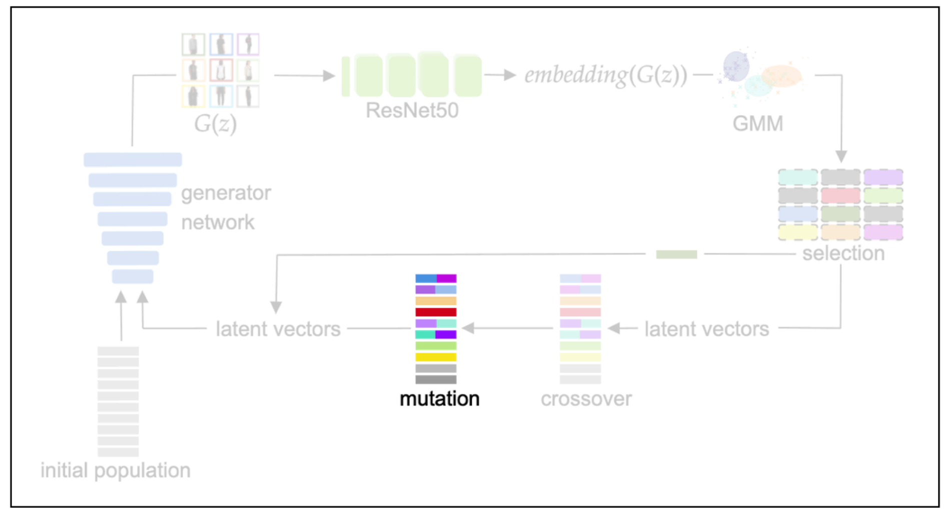
3) Genetic algorithm



3) Genetic algorithm



3) Genetic algorithm



Nonuniform mutation of individual z:

$$z_i = z_i + \alpha\, noise \sim \mathcal{N}(\mu = 0, \sigma = 1) \qquad \text{with } \alpha = Bernoulli(0.5)$$

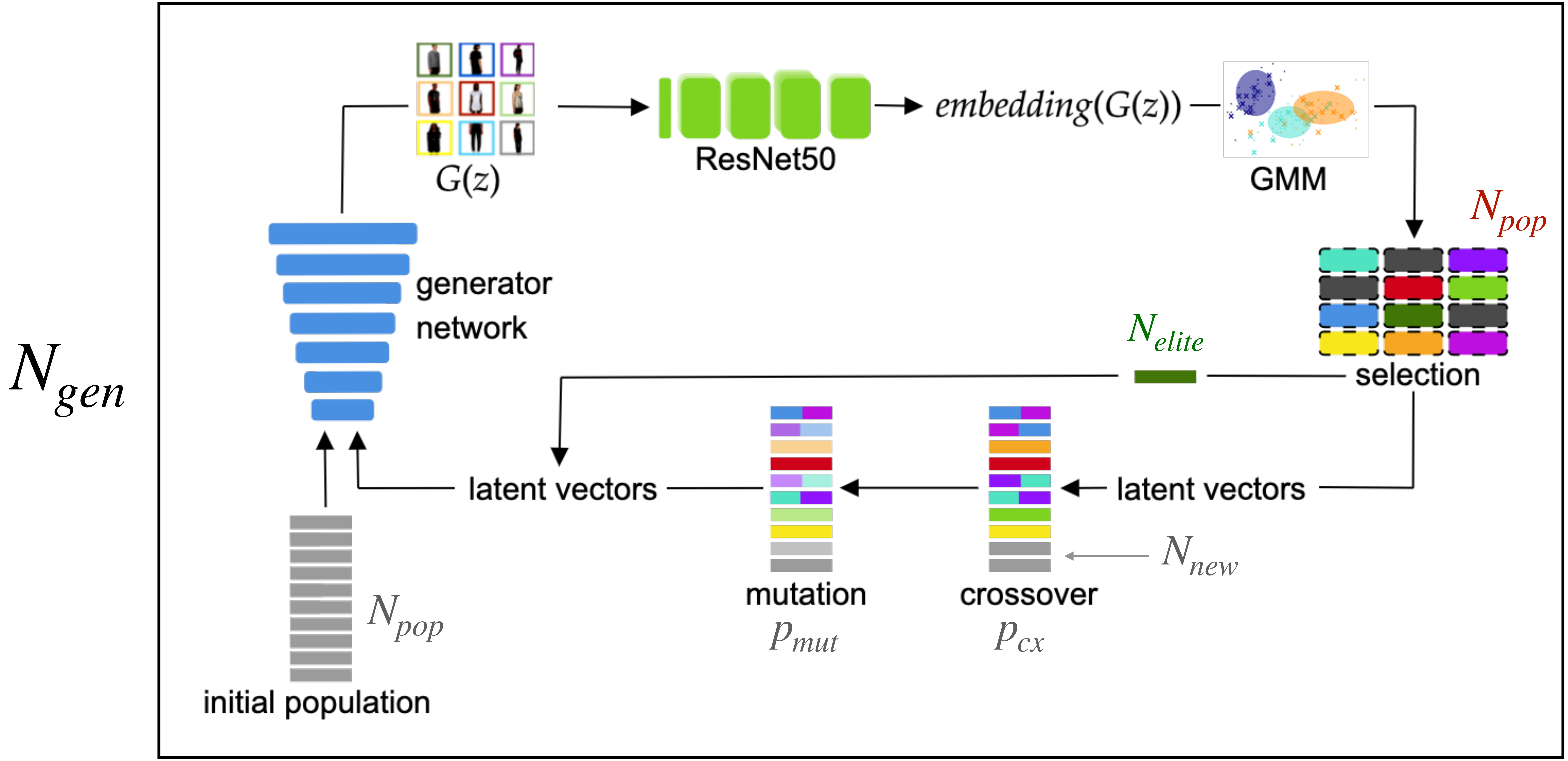
0.2	-0.6	0.9	0.4	0.1	0.5	0	-1.1	-0.1	2.3

↓

0.2	1.5	2.5	0.4	0.1	-1	0	0.3	-0.4	2.3

An individual participates in mutation by chance of mutation rate p_{mut}.

3) Genetic algorithm



Experiments

Constant GA parameters:

Parameter	Setting
Size of individual	512
$N_{generations}$	500
N_{elite}	1
N_{new}	10
Recombination operator	uniform crossover ($\alpha = 0.5$)
Mutation operator	nonuniform mutation ($\mu = 0, \sigma = 1$)

GA parameters under variation:

Parameter	Setting
Crossover rate p_{cx}	0.7, 0.9
Mutation rate p_{mut}	0.2, 0.5
Population size p_{pop}	100, 200
Tournament size p_{ts}	3, 6

Experiments

Runs for 5 style clusters per parameter combination

Style 11

Style 48

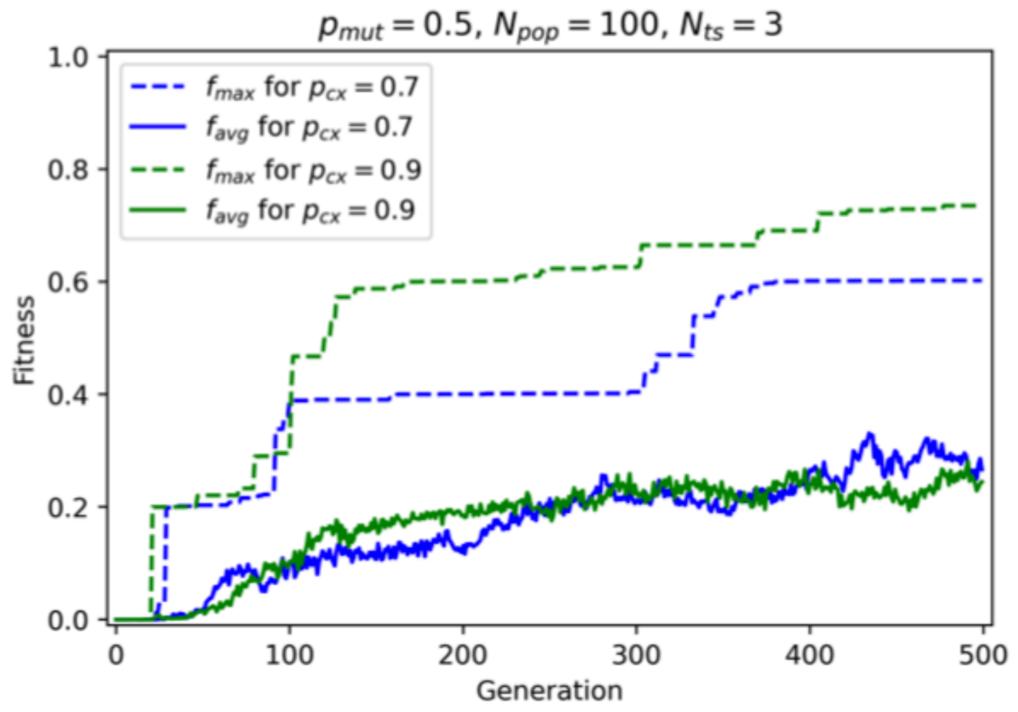
Style 65

Style 96

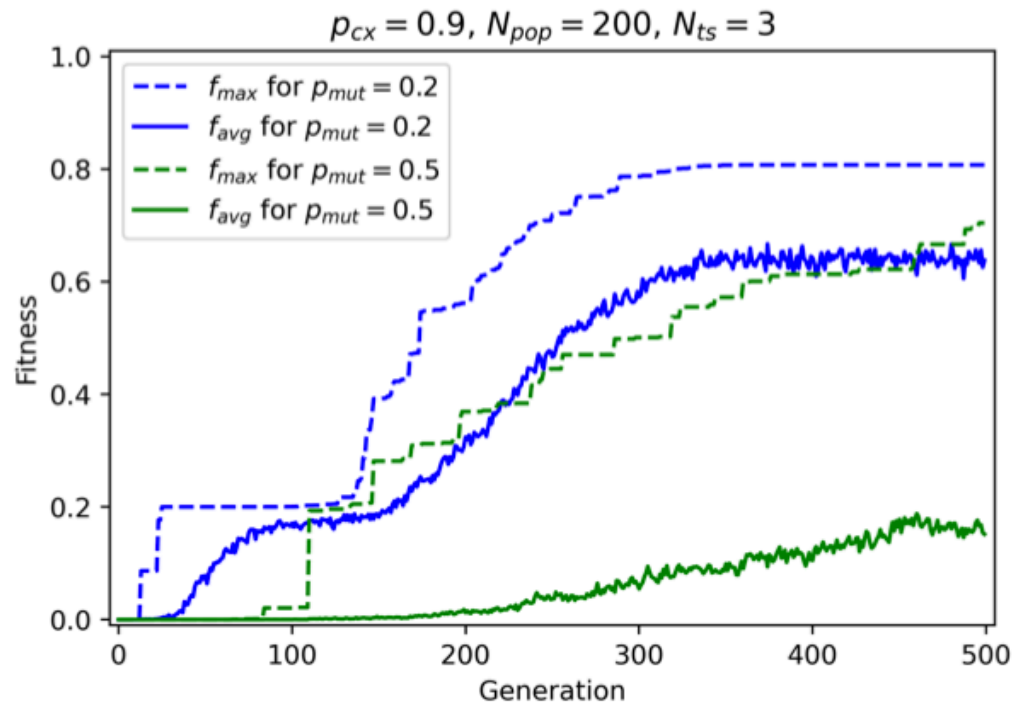
Style 138

Comparison of parameter settings

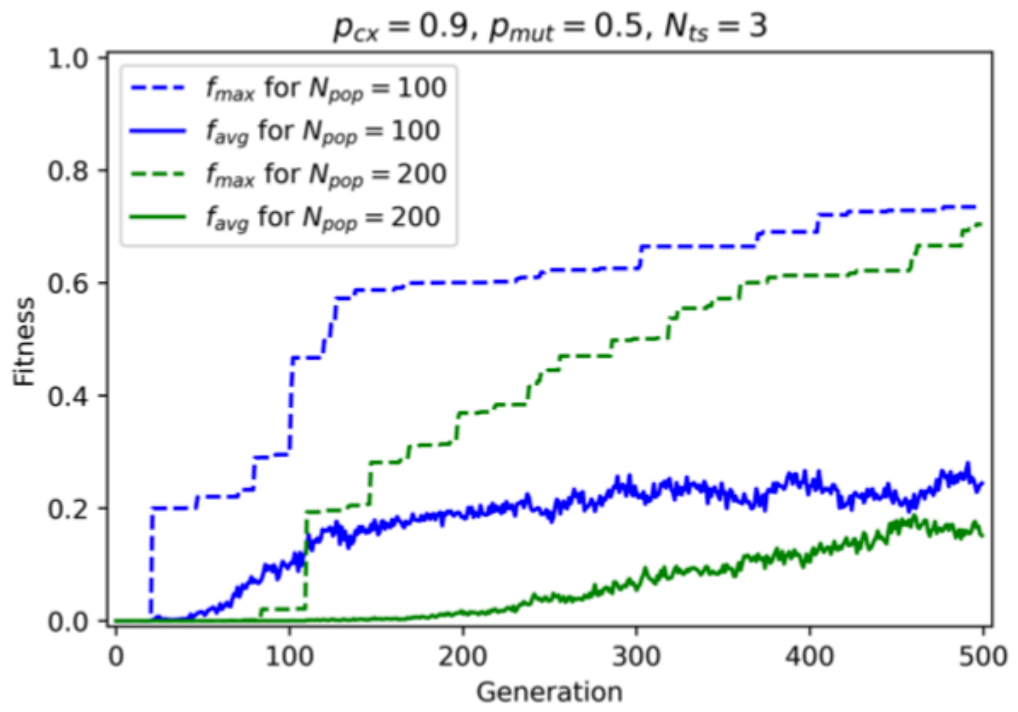
		$p_{mut} = 0.2$		$p_{mut} = 0.5$	
		$N_{te} = 3$	$N_{ts} = 6$	$N_{ts} = 3$	$N_{ts} = 6$
$p_{cx} = 0.7$	$N_{pop} = 100$	0.5746	0.2	0.6021	0.2119
	$N_{pop} = 200$	0.4031	0.5491	0.6341	0.4028
$p_{cx} = 0.9$	$N_{pop} = 100$	0.4538	0.3855	**0.735**	0.2982
	$N_{pop} = 200$	**0.8073**	0.2	0.7043	0.5248



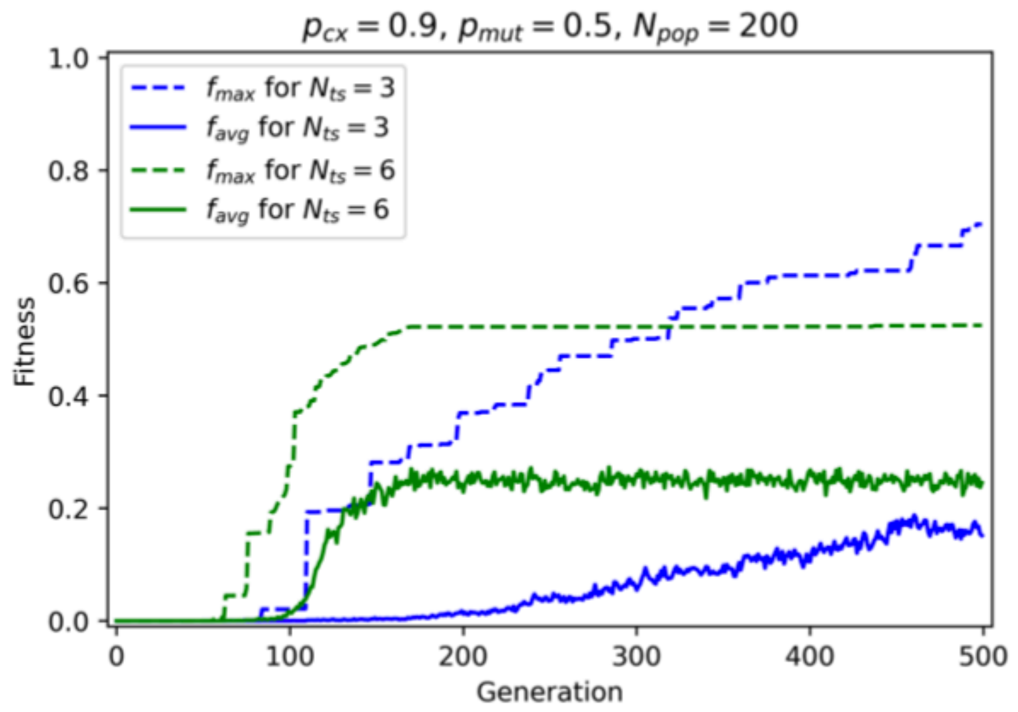
(a) Comparison of crossover rates.



(b) Comparison of mutation rates.



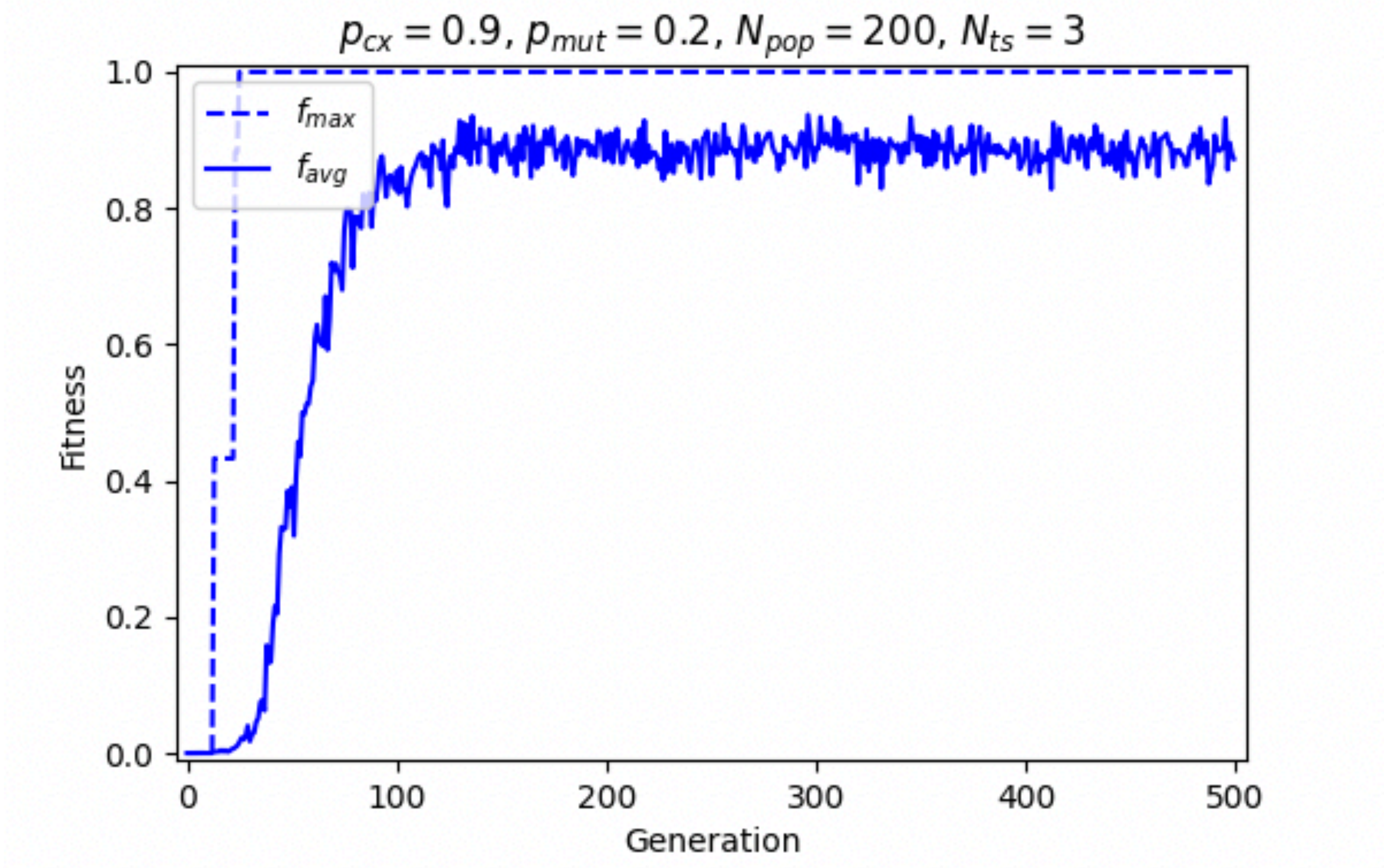
(c) Comparison of population sizes.

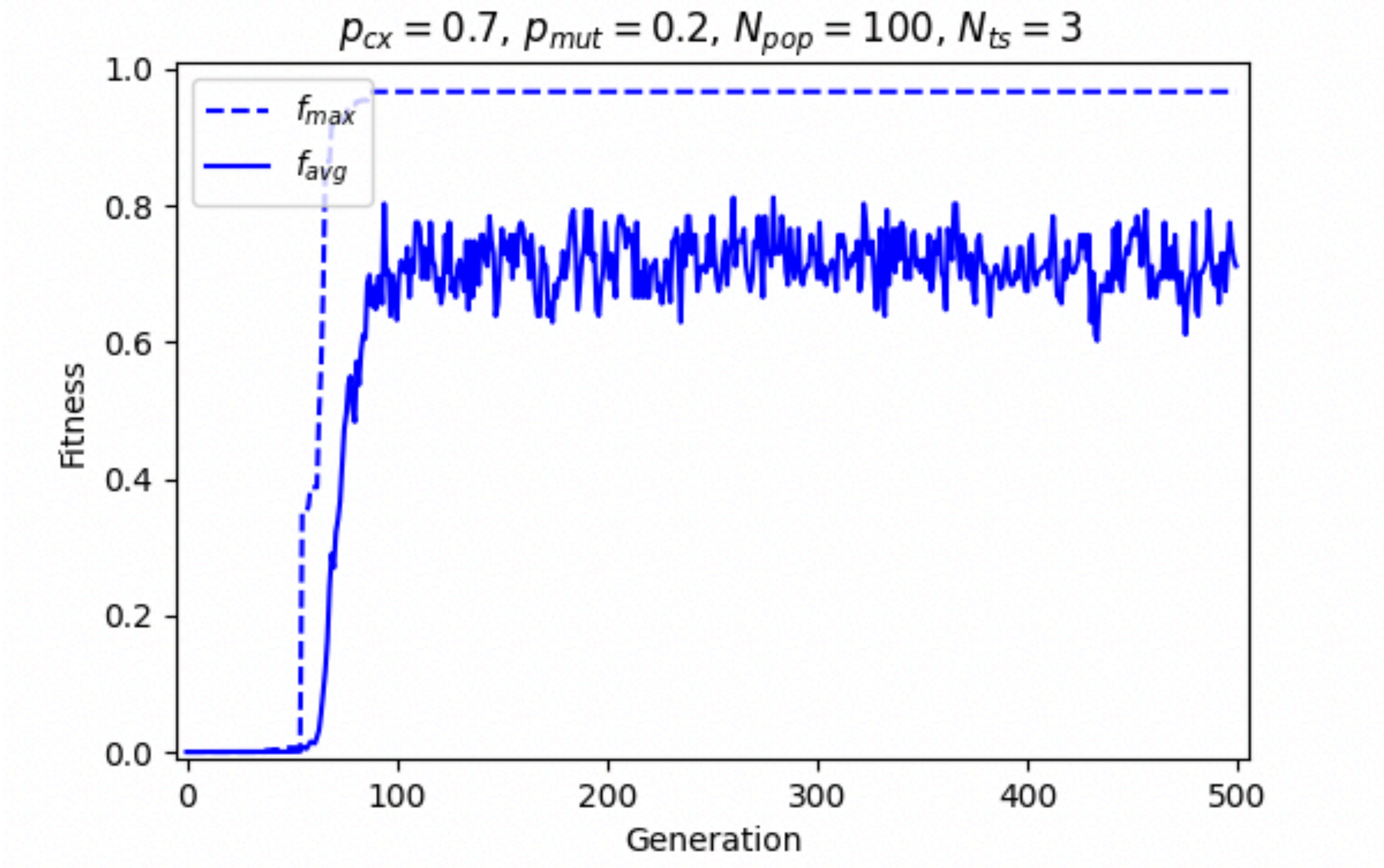


(d) Comparison of tournament sizes.

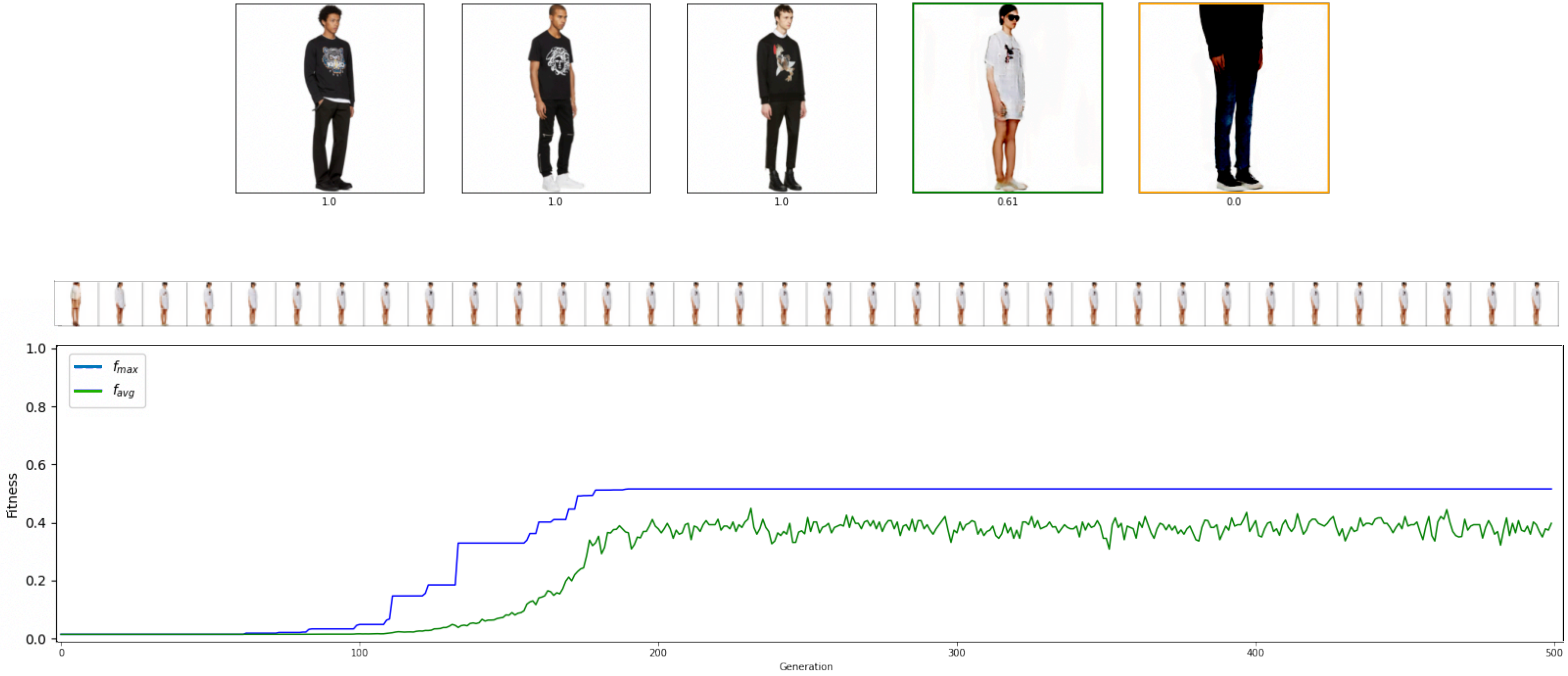
Comparison to random sampling

Fittest out if $N_{pop} * N_{gen}$ random samples

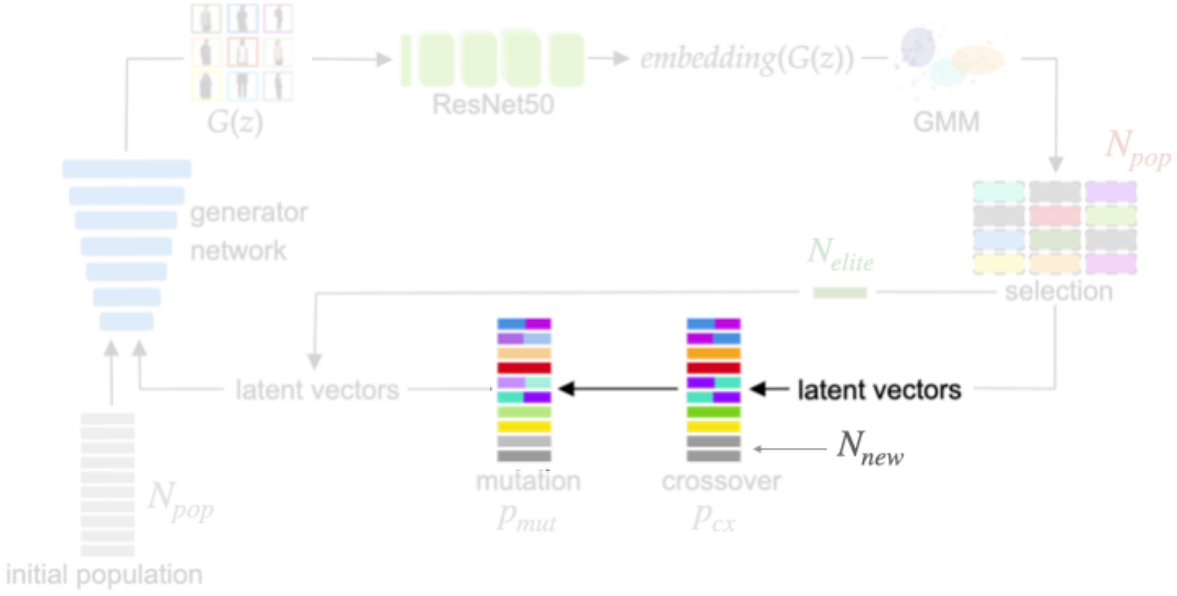


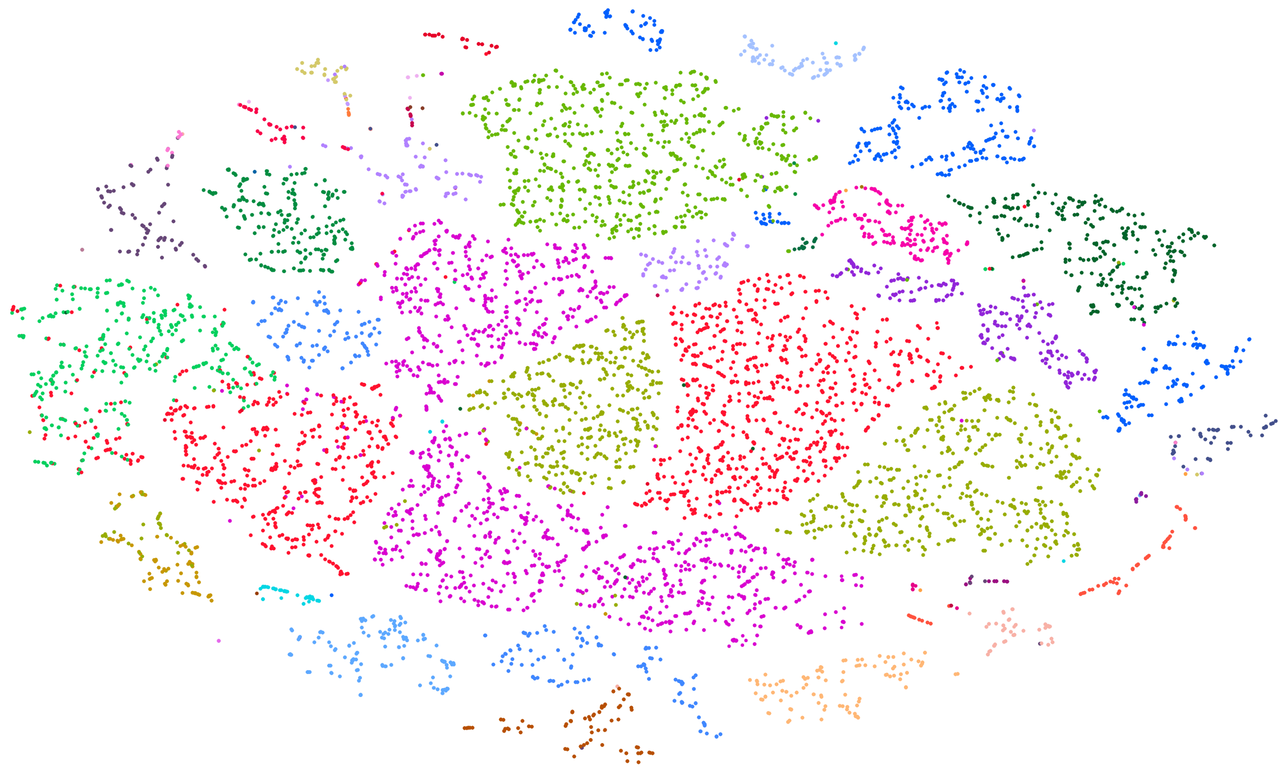


Convergence to local minima

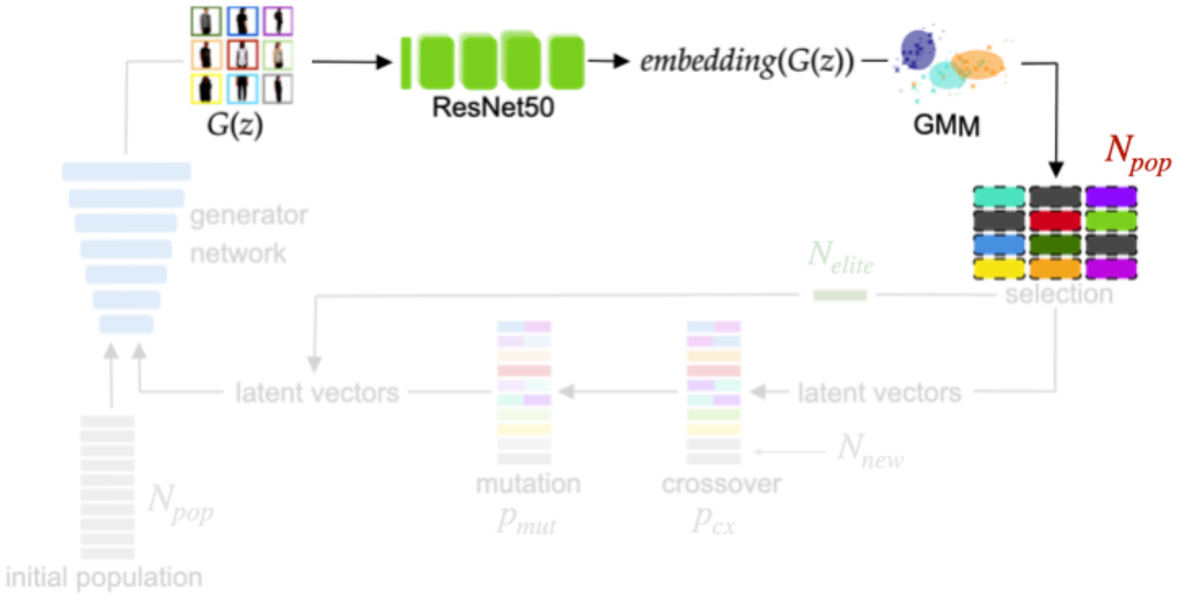


Convergence to local minima





(Rostamzadeh et al., 2018, p.7)



GA parameter setting

Visualise clusters and exploration path of population

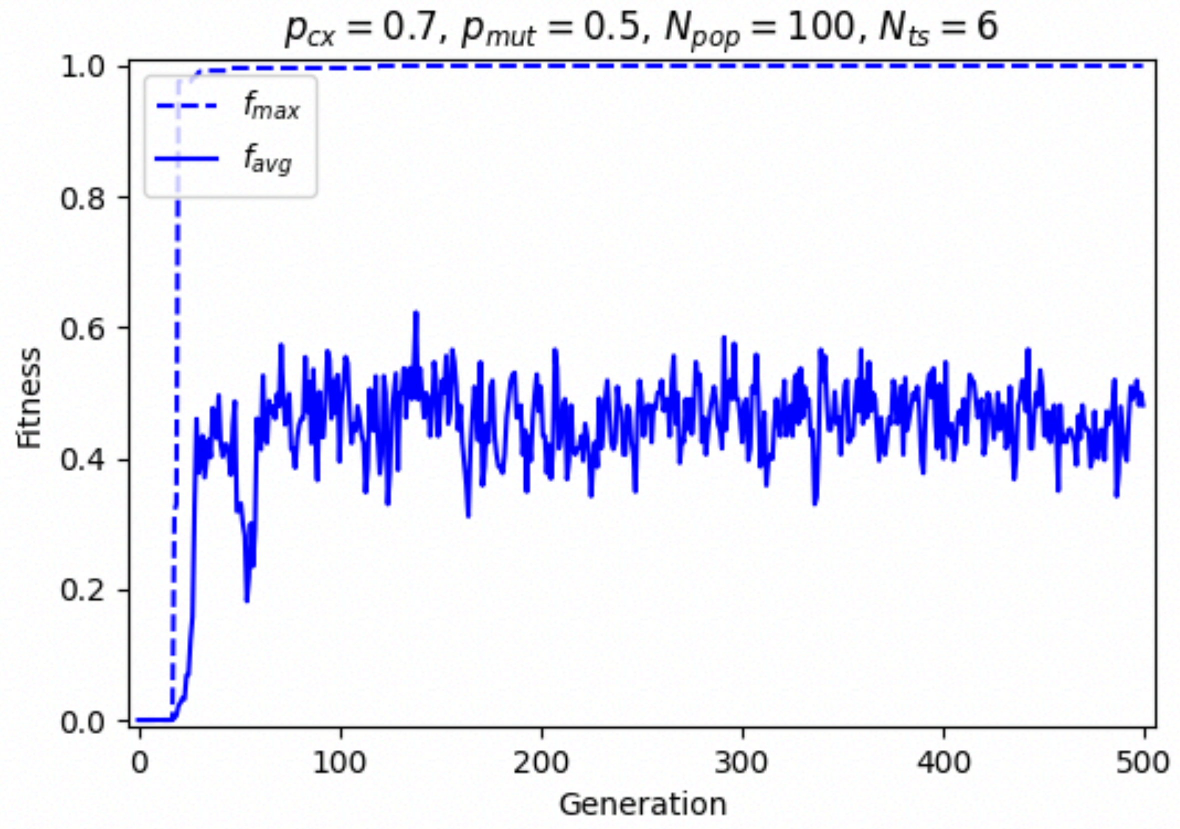
Fitness measure

What does the machine *see*?

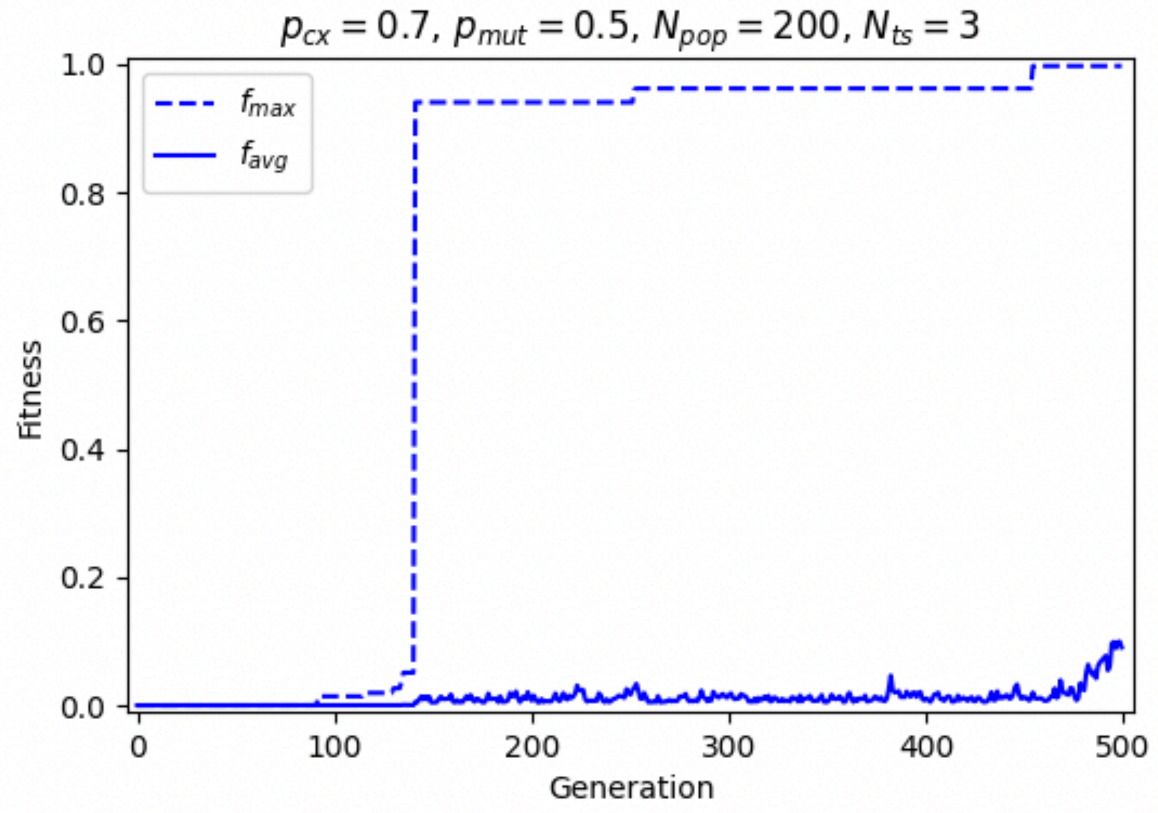

1.0


1.0


1.0



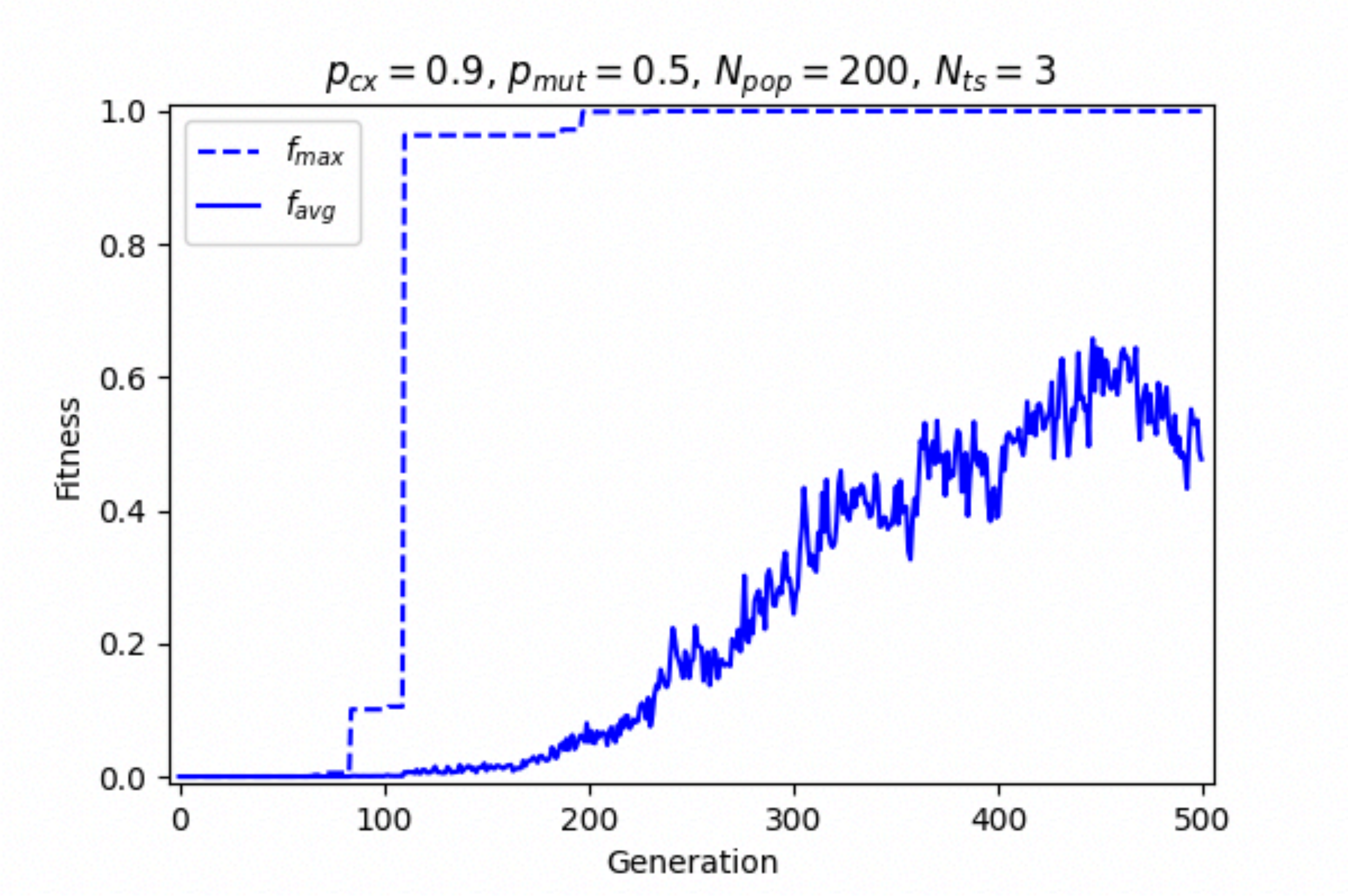

1.0




0.997

What does the machine *see*?

- Human vs. machine understanding of fashion style
- Labelled style clusters
- Interactive guidance of the evolutionary search

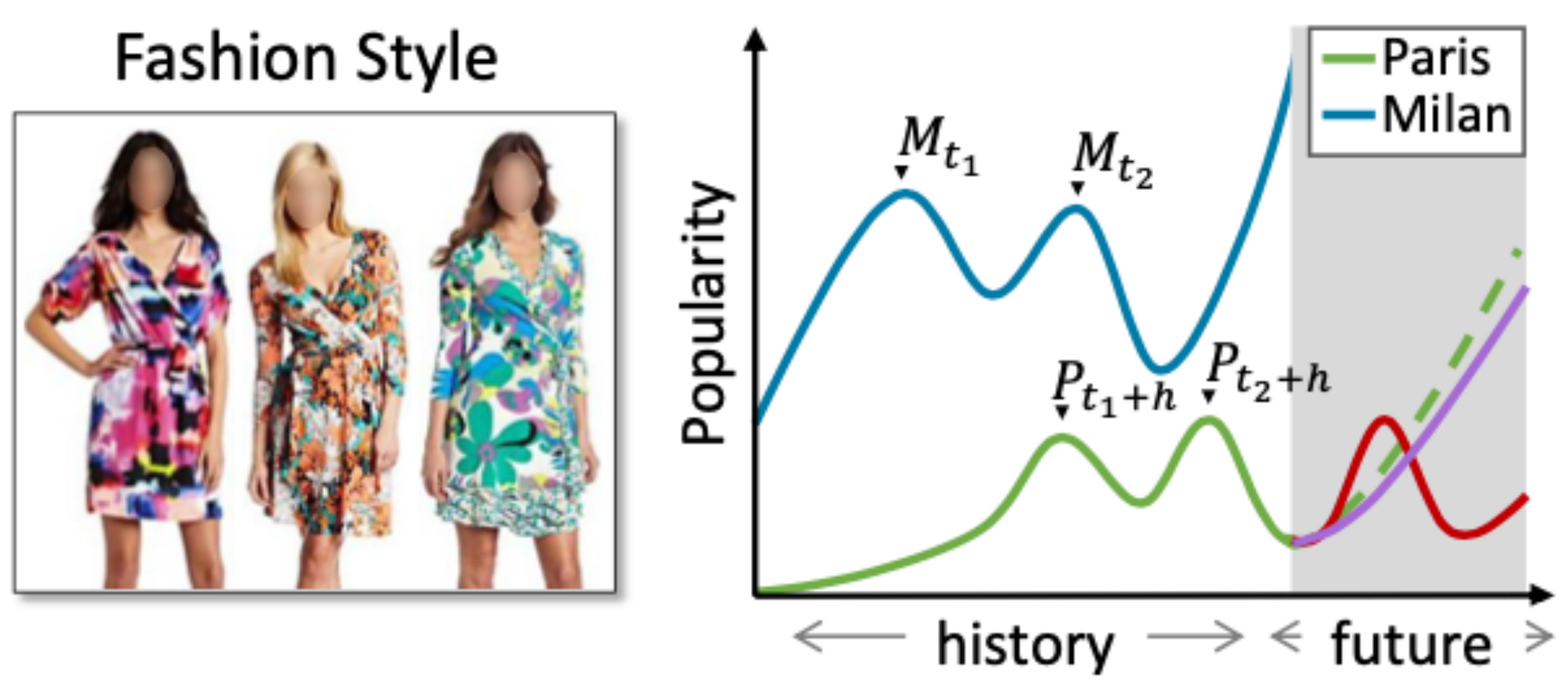


Evolutionary Search for Fashion Styles in the Latent Space of Generative Adversarial Networks

Conclusion:

- Exploration guided by fitness in accordance with style clusters with a genetic algorithm
- High fitness does not imply visual style similarity
- Improvements required to achieve reliable and stable performance

Outlook



(Al-Halah and Grauman, 2020, p.4)

Generation in response to temporal trend development

Algorithmic understanding of fashion styles

References

Al-Halah, Z. and Grauman, K. (2020). Modeling Fashion Influence from Photos. *IEEE Transactions on Multimedia*.

Bontrager, P., Lin, W., Togelius, J., and Risi, S. (2018). Deep interactive evolution. In *International Conference on Computational Intelligence in Music, Sound, Art and Design*, pages 267–282. Springer.

Fernandes, P., Correia, J., and Machado, P. (2020). Evolutionary latent space exploration of generative adversarial networks. In *International Conference on the Applications of Evolu- tionary Computation (Part of EvoStar)*, pages 595–609. Springer.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. *arXiv preprint arXiv:1406.2661*.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2017). Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv preprint arXiv:1710.10196*.

Mackinney-Valentin, M. (2010). *On the Nature of Trends: A Study of Trend Mechanisms in Contemporary Fashion*. PhD thesis.

Matzen, K., Bala, K., and Snavely, N. (2017). StreetStyle: Exploring world-wide clothing styles from millions of photos.

Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., and Pal, C. (2018). Fashion-Gen: The Generative Fashion Dataset and Challenge. *arXiv preprint arXiv:1806.08317*.

Takagi, M., Simo-Serra, E., Iizuka, S., and Ishikawa, H. (2017). What makes a Style: Experi- mental Analysis of Fashion Prediction. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2247–2253.

Yildirim, G., Jetchev, N., Vollgraf, R., and Bergmann, U. (2019). Generating high-resolution fashion model images wearing custom outfits. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0.

Zhu, S., Urtasun, R., Fidler, S., Lin, D., and Change Loy, C. (2017). Be your own Prada: Fashion Synthesis with Structural Coherence. In *Proceedings of the IEEE international con- ference on computer vision*, pages 1680–1688.

Thank you for the attention!