Appendix

A. Hierarchical Human Face Attributes

As mentioned in Sec. 4.1, a hierarchical attribute structure, which includes basic visual characteristic, is the prerequisite of PPE. In Table 2, we give a complete version of the hierarchical attribute structure for human face. The construction steps are as follows:

- Firstly, based on common sense, we establish the category-items, *e.g.*, gender and hair.
- Secondly, we use the large-scale pre-trained language model BERT [5] to mine useful human face attributes under each category. We design a series of prompts to let the pre-trained BERT predict specific attributes for every category, using an open-source tool, *i.e.*, LAMA [34]. We find prompts like *"a face/person with a/an [MASK] [X]"*, *"the person's [X] is/are [MASK]"* and *"a [MASK] [X]"* are helpful, where *[MASK]* is the word that BERT has to predict and *[X]* is the concrete category (*e.g.*, hair, nose, mouth and etc.) in practice.
- Furthermore, among the many predicted attributes, we select the representative ones and sort them into more fine-grained categories, such as hair color and hair length.
- Lastly, we include some binary attributes, especially drawing on the previously proposed human face attributes in the CelebA dataset [25].

B. Category Finder NLP Tool

Finding the corresponding category of command is a step in PPE. For example, we should find that the category of *"male"* is *gender* and the category of *"black hair"* is *hair color*. Thereby, we can exclude the attributes of the same category as the command, then predict the entangled attributes from the rest (Sec.4.1).

The finding operation can be performed automatically or manually. In this paper, since we conduct experiments mainly on single attributes, which means there will be only one category of the command, we can use a simple NLP tool based on pre-trained language models. Specifically, we design a prompt, *i.e.*, *"[Y] is a kind of [X]"*, where *[Y]* is the input command text and *[X]* is the candidate category. For each candidate category, we fill the prompt with the given command text and the category, then compute the perplexity of the sentence using a pre-trained BERT (through LAMA). The category with the lowest perplexity will be suggested as the most possible one. The method is proven effective.

C. Generability

We extend our method to the multiple attribute manipulations by handling them as multiple single-command, predicting entangled attributes for each one, and using Entanglement Loss based on the grouped predicted entangled attributes. The experimental results are illustrated in Fig. 8. As shown, our method can be generalized to the multiple attribute manipulation setting.

Category	Attributes
gender	male, female
hair color	black hair, blond hair, brown hair, grey hair, red hair
hair length	long hair, short hair, no hair
hair style	curly hair, straight hair, bald, wavy hair, receding hairline
eye color	blue eyes, brown eyes, black eyes, grey eyes, green eyes
eye status	open eyes, closed eyes
eye shape	narrow eyes, wide eyes, big eyes, small eyes, round eyes
nose shape	big nose, long nose,pointed nose, small nose, hooked nose, short nose, thick nose, thin nose, pinched nose, flat nose
face shape	pointy face, round face, square face, oval face, long face
skin color	white skin, black skin, yellow skin
mouth status	open mouth, close mouth
mouth size	big mouth, small mouth
eyebrows	round eyebrows, high eyebrows, arched eyebrows, long eyebrows, thick eyebrows, dark eyebrows, straight eyebrows, thin eyebrows, short eyebrows
beard	goatee, mustache, no beard, sideburns, 5 o'clock shadow
earrings	with/without earrings
makeup	with/without makeup
smile	with/without smile
lipstick	with/without lipstick
wrinkles	with/without wrinkles
glasses	with/without glasses
bangs	with/without bangs
rosy cheeks	with/without rosy cheeks
bags under eyes	with/without bags under eyes
high cheekbones	with/without high cheekbones

Table 2. The complete hierarchical human face attributes.

Figure 8. Multiple attribute manipulations.