

## Image-to-Text Inversion User Studies

This questionnaire survey focuses primarily on the evaluation of editing effects combining images and text. Each question provides a reference image and the textual form of the editing target, with 1-2 editing targets highlighted in red. Participants are required to consider the following three aspects (in order of importance): 1. The degree of conformity between the result and the editing statement. 2. The degree of similarity between the result and the reference image (only for the foreground, such as the facial area in questions 1-10, excluding hairstyles). 3. The overall quality and naturalness of the image. Participants are to select the best editing result for each question, and each question is a single-choice format.

1. Editing target: A woman with **teal hair**, **in front of a pink wall**. Reference :



○1



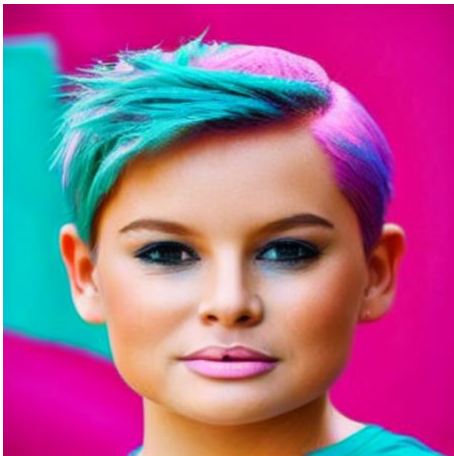
○2



03



04



05



06



07

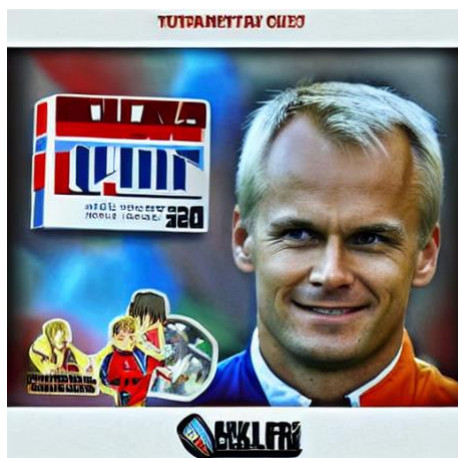


08



o9

2. Editing target: A man face funko pop. Reference :



o1



o2





o3



o4



o5



o6



o7



o8



o9

3. Editing target: An **oil painting** of a woman face **in a wooden frame**. Reference :



o1



o2



o3



o4



o5



o6



o7



o8





o9

4. Editing target: A **tattoo** of a man's face **on the back**. Reference :



o1



o2



03



04



05



06

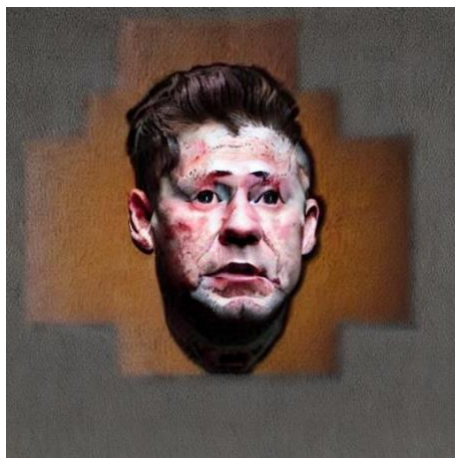


07



08



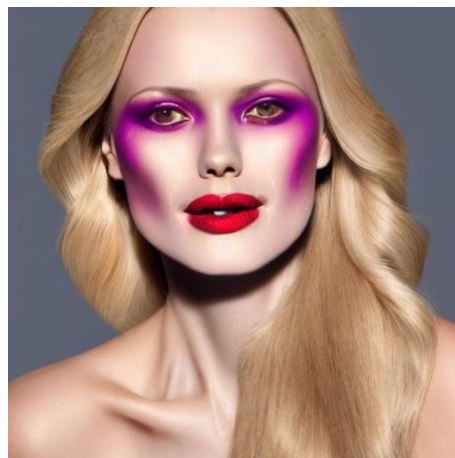


o9

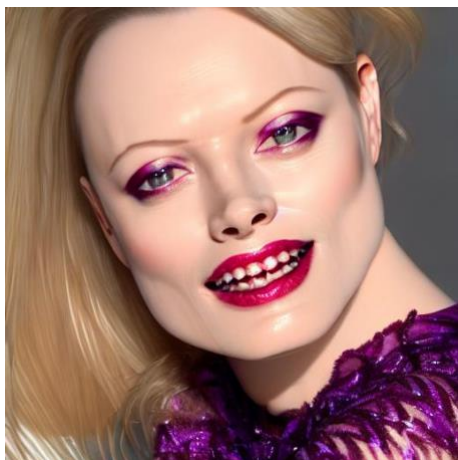
5. Editing target: A woman with **red lipstick** and **purple eye shadow**. Reference :



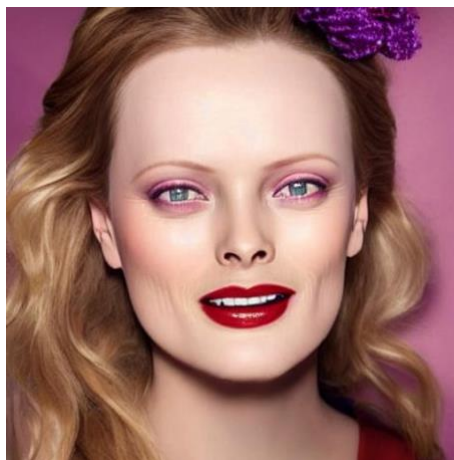
o1



o2



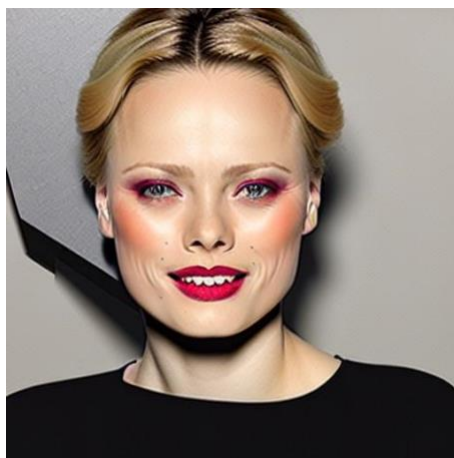
o3



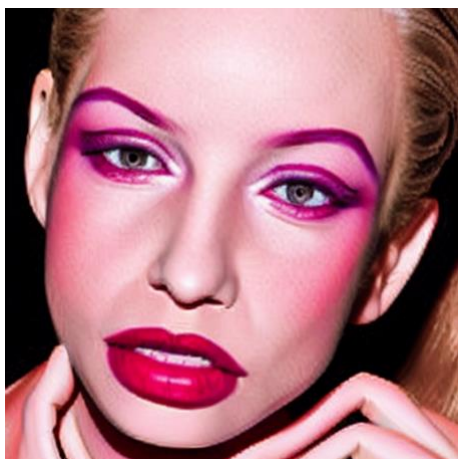
o4



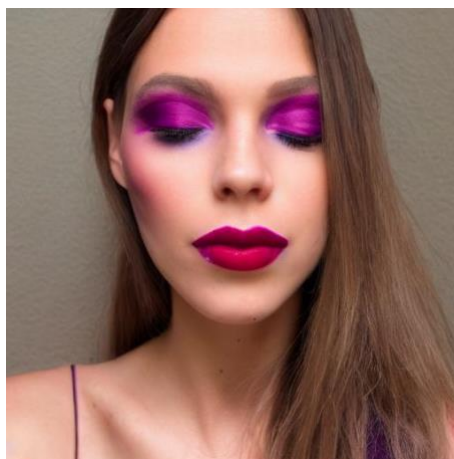
o5



o6



o7



o8



o9

6. Editing target: A man with **long hair**, **side view**. Reference :



o1



o2





o3



o4



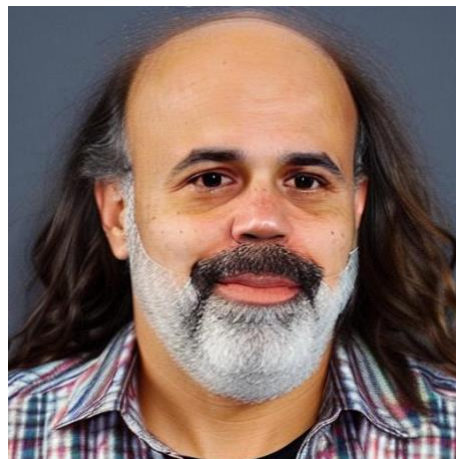
o5



o6



o7



o8



o9

7. Editing target: A woman wearing **nurse uniform**. Reference :



o1



o2



o3



o4



o5



o6

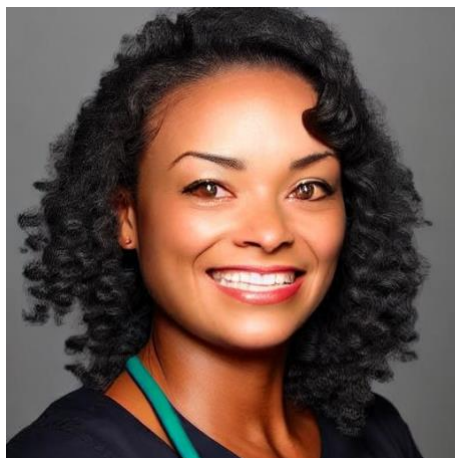


o7



o8





o9

8. Editing target: A man **face on a metal coin**. Reference :



o1



o2



o3



o4



o5



o6



o7



o8



o9

9. Editing target: A woman **as a WNBA player**. Reference :



o1



o2

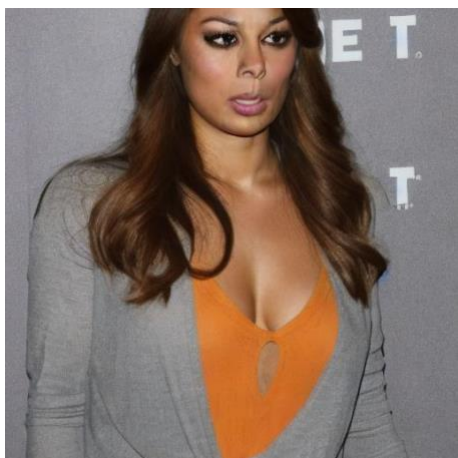




o3



o4



o5



o6



o7

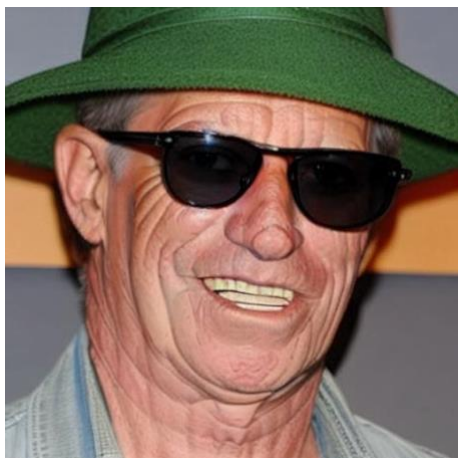


o8

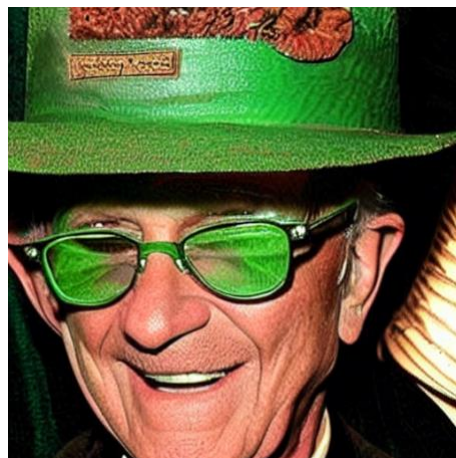


o9

10. Editing target: A man **wearing sunglasses** and **a green helmet**. Reference :

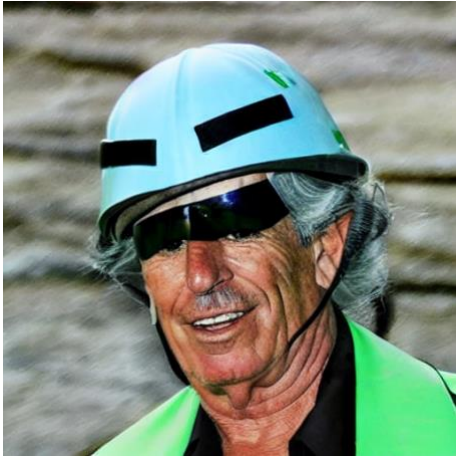


o1



o2





o3



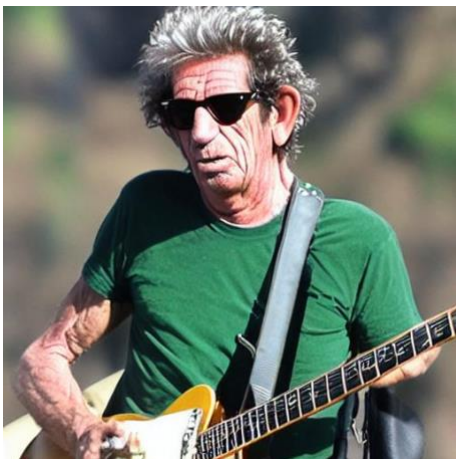
o4



o5



o6



o7



o8





o9

11. Editing target: A toy **on the sofa**. Reference :



o1



o2



o3



o4



o5



o6



o7



o8

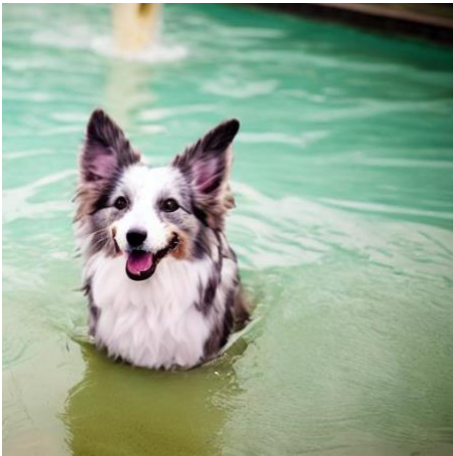
12. Editing target: A dog **swimming in the pool**. Reference :



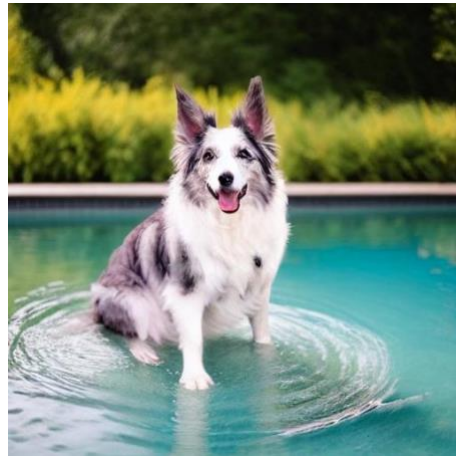
o1



o2



o3



o4





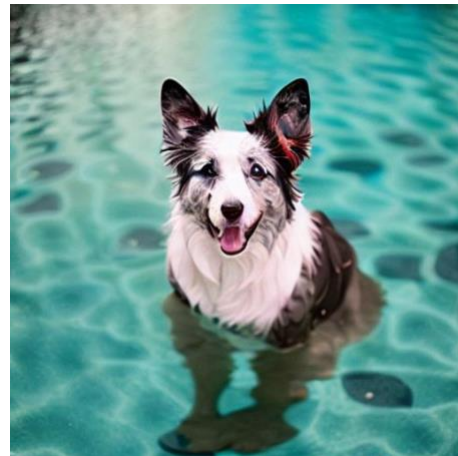
o5



o6



o7



o8

13. Editing target: **A pair of boots.** Reference :





o1



o2



o3



o4



o5



o6

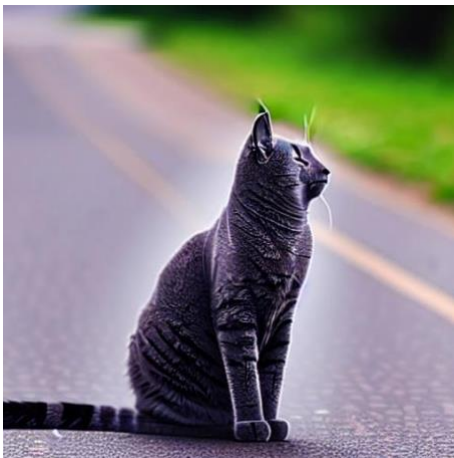


o7



o8

14. Editing target: A cat **walking on the road**. Reference :



o1



o2





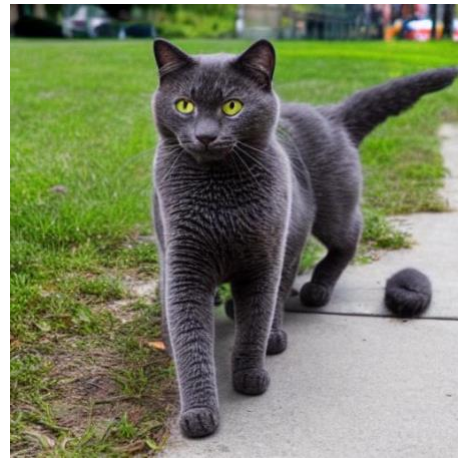
o3



o4



o5



o6



o7

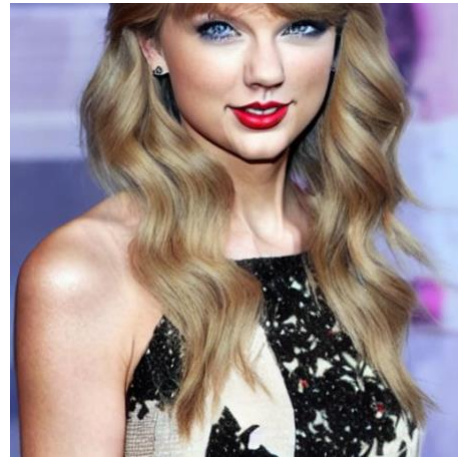


o8

15. Editing target: **Taylor Swift** with **this hair**. Reference :



o1



o2



o3



o4



o5



o6



o7



o8

16. Editing target: Gal Gadot wearing that clothes. Reference :







o1



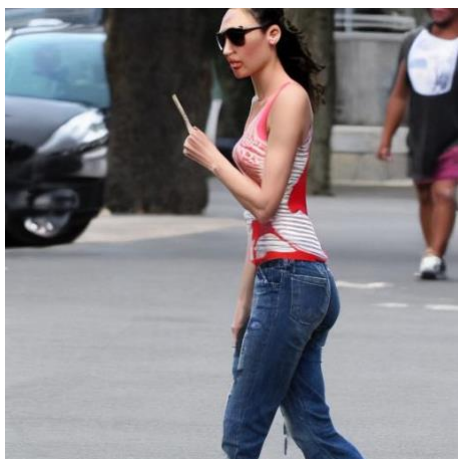
o2



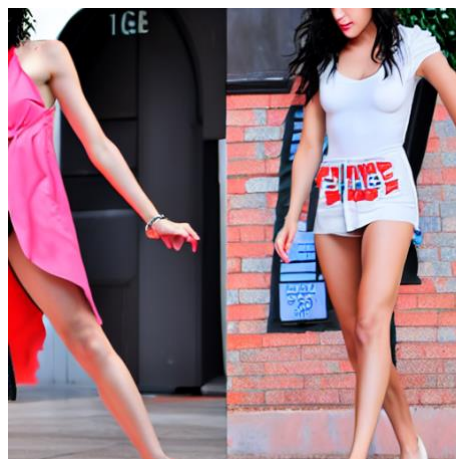
o3



o4



o5



o6

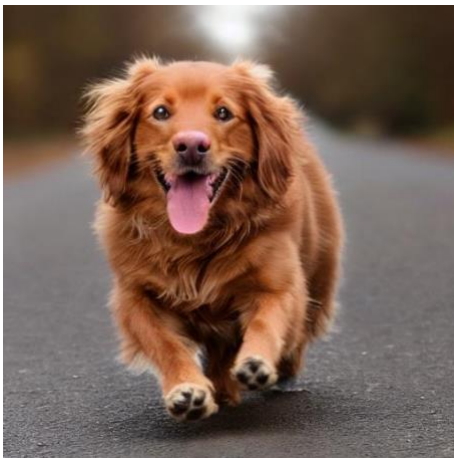


o7

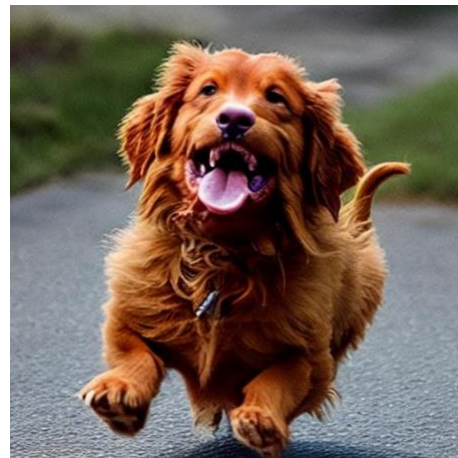


o8

17. Editing target: A dog **running on the road**. Reference :



o1

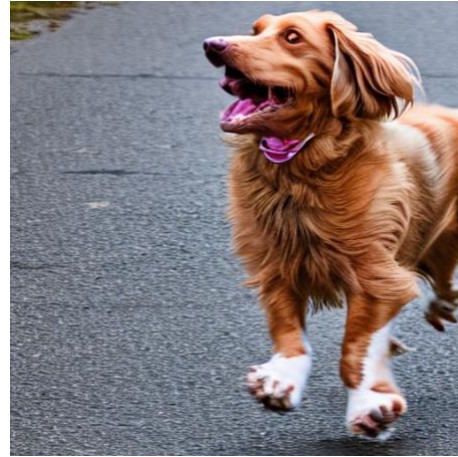


o2

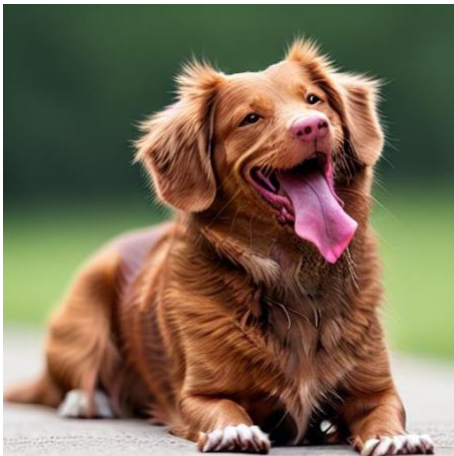




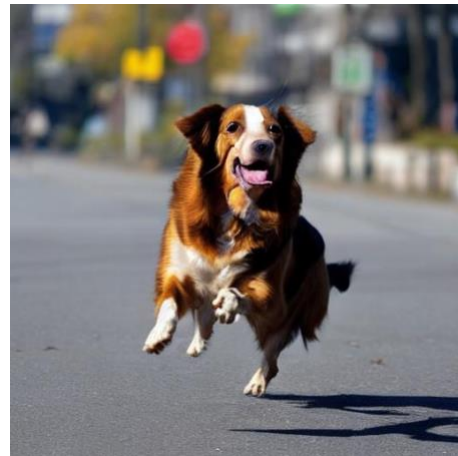
o3



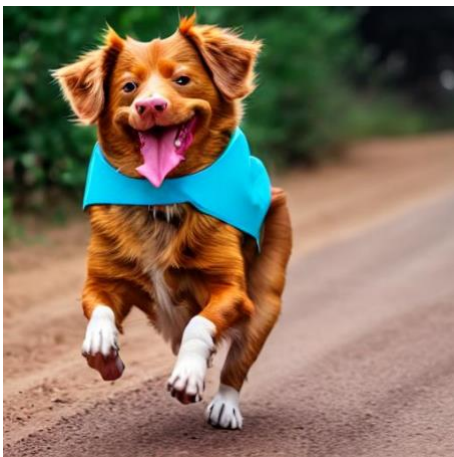
o4



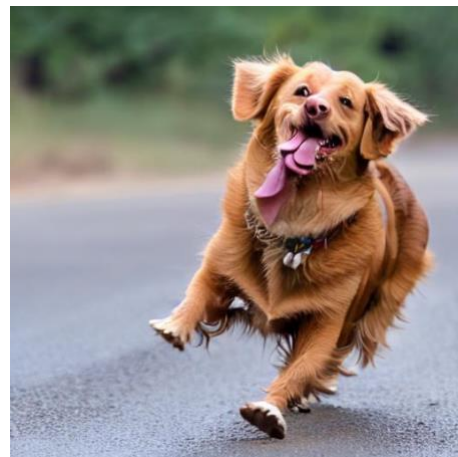
o5



o6



o7



o8



18. Editing target: A vase **on the beach** with **roses in it**. Reference :



o1



o2



o3



o4



o5



o6



o7



o8

19. Editing target: A girl **wearing that backpack**. Reference :





o1



o2



o3



o4



o5



o6





o7



o8

20. Editing target: A clock **on the table showing twelve o'clock**. Reference :



o1



o2



o3



o4



o5



o6



o7



o8

