





Text Ground-Truth Objects



Th

Shikra

Regular Decoding

The image captures a nighttime view of a city skyline with a lit-up clock tower. The clock face is visible, displaying the **time as 10:45**. The city skyline is bustling, with numerous buildings and a few clocks on display. In the foreground, a **car** can be seen, likely moving through the busy city streets. The scene creates an impression of a vibrant urban area, where **people** are enjoying their time and engaging with the city's architecture and nightlife.



1

Shikra

EAZY

The image captures a nighttime cityscape shot from a high vantage point, showcasing the skyline with tall buildings lit up. There is a prominent clock tower in the scene, situated on the left edge, with a smaller clock visible closer to the center. The cityscape is illuminated by the glow of the night, highlighting the bright lights on the buildings. Overall, the scene creates a lively and dynamic atmosphere.