Deepfakes I

IMPORTANT NOTES:
Study lecture materials at least 1 hour and prepare the questions prior to the tutorial session.
The questions will be discussed in the tutorial session.

1. What are the three core security properties discussed in the lecture, and how can AI compromise each of them?

   - Confidentiality (CONF): Ensures that information is not disclosed to unauthorized individuals. AI can compromise this through inference attacks, where it deduces sensitive information from seemingly anonymized data.
   - Integrity (INT): Ensures that data remains unaltered from its original source. AI can generate deepfakes that tamper with images, videos, or audio without detection.
   - Authentication (AUTH): Verifies the identity of users or sources. AI can mimic biometric traits (e.g., facial expressions, voice) to impersonate individuals, undermining authentication systems.

2. Explain the difference between encryption and inference attacks in the context of confidentiality. How does AI enhance the threat of inference attacks?

   Encryption transforms data into an unreadable format to prevent unauthorized access. Inference attacks exploit patterns in data to deduce sensitive information without direct access. AI enhances inference attacks by using machine learning models to recognize patterns, predict relationships, and cluster similar data, making it possible to reverse-engineer private information from anonymized datasets.

3. What is the role of keypoint detection in the First Order Motion Model for image animation? Why is it critical for generating realistic deepfakes?

   Keypoint detection identifies landmarks (e.g., eyes, mouth) in both source and driving images. These keypoints guide the motion transfer, ensuring that the source image mimics the movements of the driving video. Accurate keypoint detection is critical because it preserves facial structure and expression, making the deepfake appear realistic and coherent.

4. Describe the brightness constancy assumption in optical flow. Why is this assumption important for motion estimation in deepfake generation?

   The brightness constancy assumption in optical flow assumes that the intensity of a pixel remains constant as it moves across frames. This enables accurate motion tracking by comparing pixel displacement. In deepfakes, this helps maintain visual consistency during animation, ensuring smooth and believable transitions.

5. In motion-supervised co-part segmentation, how is motion used to identify and segment object parts? What advantages does this self-supervised approach offer over traditional supervised methods?

   Motion is used to group pixels that move together, forming segments. The model learns from pairs of frames without labeled data.

   The advantages are as follow:

   - Reduces reliance on annotated datasets.
   - Adapts to diverse appearances.
   - Enables scalable and flexible segmentation for deepfake generation.

6. Compare affine and projective warping transformations. How do these affect the realism and accuracy of deepfake animations?

   Affine transformations include scaling, rotation, translation, and shearing. They preserve parallel lines and are simpler. Projective transformations allow for perspective distortion, enabling more complex and realistic warping. Projective warping enhances realism by simulating depth and viewpoint changes, crucial for high-quality deepfakes.

7. A company uses facial recognition for employee authentication. A deepfake video mimics an employee's facial gestures to gain unauthorized access. What type of biometric authentication is being attacked, and how could the system be improved to resist such deepfake threats?

   The attack targets soft biometrics (e.g., facial gestures, expressions). Improvements are as follow:

   - Use multi-factor authentication.
   - Incorporate liveness detection (e.g., blinking, depth sensing).
   - Employ dynamic biometrics like gait or keystroke patterns.