


Cuestionario 3 - VC

Óscar David López Arcos. 

1. ¿Cuáles son las propiedades esenciales que permiten que los modelos de recuperación de instancias de objetos de una gran base de datos a partir de descriptores sean útiles? Justificar la respuesta.

Si son realmente representativos, los descriptores de distintas imágenes de una misma instancia serán bastante similares. Esto se debe a las propiedades de los descriptores, invariables a las transformaciones de intensidad, rotación y escala.

Si tenemos una gran base de datos, podemos agrupar los descriptores en clusters calculando el centroide de estos para que posteriormente, al evaluar una nueva imagen, sea sencillo dictaminar a qué instancia pertenece de forma simple y eficiente.

2. ¿Justifique el uso del modelo de bolsa de palabras en el proceso de detección y reconocimiento de instancias de objetos? ¿Qué ganamos? ¿Qué perdemos? Justificar la respuesta.

Con el modelo de bolsa de palabras pasamos de calcular las correspondencias entre los descriptores a calcular correspondencias entre las palabras de las imágenes. Cada palabra se define como la representación de varios descriptores que se encuentran en el mismo cluster.

De esta forma, ganamos en representación. Al establecer este “margen” para los descriptores, dos imágenes que representen la misma instancia con pequeñas transformaciones debidas a la perspectiva, podrán estar representadas por las mismas palabras aunque no lo estén por los mismos descriptores.

Por otro lado, perdemos a la hora de diseñar el algoritmo, ya que establecer el tamaño de los clusters que llevarán cada descriptor a una única palabra es un proceso complicado. Un tamaño demasiado grande llevará características distintas a la misma palabra y un tamaño pequeño podrá concluir en la misma característica representada por dos palabras diferentes.

3. ¿Describa la diferencia esencial entre los problemas de reconocimiento de instancias y reconocimiento de categorías? ¿Qué deformaciones se presentan en uno y otro? Justificar la respuesta.

Como su propio nombre indica, el reconocimiento de categorías es capaz de reconocer objetos distintos que formen parte de una misma categoría (por ejemplo, reconocer gatos) mientras que el reconocimiento de instancias se entrena para reconocer un único y determinado objeto (siguiendo el mismo ejemplo, reconocer un gato en concreto).

En ambos problemas podremos encontrar deformaciones geométricas y proyectivas, pero al problema de reconocimiento de categorías se le añade la dificultad de que los objetos dentro de una misma categoría pueden tener muchas características no comunes. Por eso es tan importante la correcta selección de características en el entrenamiento del modelo.

4. ¿Es posible usar el modelo de bolsa de palabras para el reconocimiento de categorías de objetos? Justificar la contestación.

Solamente podría aplicarse este modelo para el reconocimiento de categorías en entornos muy controlados. Por ejemplo, en el reconocimiento de sillas visto en las diapositivas, podríamos aplicarlo siempre y cuando todas las sillas aun siendo distintas fuesen muy parecidas entre sí,

presentando colores y formas similares. En este caso, los descriptores serán similares por lo que el reconocimiento basado en bolsa de palabras podría funcionar. En otro caso, como ya comentamos anteriormente, sería totalmente imposible.

5. Suponga que desea detectar, en una imagen, una instancia de un objeto a partir de una foto del mismo tomada desde el mismo punto de vista del que aparece en la imagen y en un entorno de iluminación similar. Analice la situación en el contexto de las técnicas de reconocimiento de objetos e identifique que algoritmo concreto aplicaría que fuese útil para cualquier objeto. Argumente porqué funcionaría y especifique los detalles necesarios que permitan entender su funcionamiento.

Siendo el mismo punto de vista y presentando condiciones lumínicas similares, la forma más rápida de identificar un objeto sería mediante el algoritmo de ventana deslizante. Al basarse en la intensidad de gradiente, si elegimos un tamaño de ventana adecuado, responde con mayor intensidad en el punto que detecte el objeto.

Aunque este procedimiento tiene sus contras, como la posibilidad de generar falsos positivos, en las circunstancias descritas en el enunciado debería funcionar bastante bien.

6. Suponga de nuevo el problema del ejercicio anterior pero la foto que le dan está tomada con un punto de vista del objeto distinto respecto del objeto en la imagen. Analice que repercusiones introduce esta modificación en su solución anterior y que cambios debería de hacer para volver a tener un nuevo algoritmo exitoso. Justificar la respuesta.

En este caso, al tratarse de reconocer el mismo objeto, podríamos aplicar un algoritmo de detección de instancias como podría ser una bolsa de palabras.

Respecto a la solución anterior este método es mucho más complejo, obligándonos a definir un vocabulario e histogramas para cada imagen, para así poder identificar si se trata del objeto en cuestión en función del número de palabras que tengan en común.

7. Suponga que una empresa de Granada le pide implementar un modelo de recuperación de información de edificios históricos de la ciudad a partir de fotos de los mismos. Explique de forma breve y clara que enfoque le daría al problema. Que solución les propondría. Y como puede garantizar que la solución podrá ser usada de forma eficiente a través de dispositivos móviles.

Una vez más, al ser reconocimiento de instancias, el modelo bolsa de palabras sería una buena solución. Sin embargo, en el caso particular de los edificios, a pesar de ser distintos pueden compartir palabras comunes (ya que la estructura siempre será similar). Para paliar este problema, podemos aplicar RANSAC entre las correspondencias de la imagen-pregunta y todas las posibles respuestas. Aquella imagen que mantenga mayor número de correspondencias tras aplicar RANSAC será el edificio al que nos referimos.

Para que pueda realizarse en un móvil de forma eficiente, el usuario podría tomar la foto y enviarla a un servidor. Este calcularía los descriptores y palabras asociadas a esta y la compararía con el modelo de índice invertido ya calculado previamente.

8. Suponga que desea detectar la presencia/ausencia de señales de tráfico en imágenes tomadas desde una cámara situada en la parte frontal de un coche que viaja por una carretera. Diga que aproximación usaría y porqué. Identifique las principales dificultades y diga como las resolvería. Los argumentos deben ser sólidos y con fundamento en las técnicas estudiadas.

El problema de detección de señales de tráfico es fácilmente tratable por un modelo basado en ventana de reconocimiento genérico de categorías. El modelo puede entrenarse usando muchas imágenes de señales, definiendo las características que creamos representativas.

Estos algoritmos funcionan mejor para objetos rígidos con una estructura 2D firme, como es el caso de las señales de tráfico.

Uno de los principales problemas de este algoritmo sería la gran complejidad computacional que presenta. Para solucionarlo, se podría intentar entrenar para la detección de no-señales y, de forma similar a como Viola-Jones hace con las caras, descartar de forma prematura aquellas imágenes cuando existe certeza suficiente de que no aparecen señales.

9. ¿Qué han aportado los modelos CNN respecto de los modelos de reconocimiento de objetos empleados hasta 2012? Enumerar las propiedades comunes entre ellos y aquellas claramente distintas que hayan permitido una mejora en la solución del problema por parte de las CNN. Dar una opinión razonada de por qué significan realmente una mejora.

Desde 2012 en adelante, con la exitosa implementación del Deep Learning la obtención de características ha mejorado sustancialmente, demostrando que con unas buenas características hasta el clasificador más simple puede resolver un problema complejo.

Propiedades comunes:

- Ambos necesitan gran cantidad de imágenes para funcionar correctamente.
- Ambos trabajan con características obtenidas de las imágenes para compararlas.
- Tienen un entrenamiento costoso.

Propiedades distintas:

- Los modelos CNN extraen información de las imágenes para obtener características, mientras que en los métodos tradicionales eran escogidas por el diseñador.

Como conclusión, hoy día, con la cantidad de información fácilmente accesible y las capacidades de cómputo, las CNN ofrecen respuestas a problemas inabordables hasta el momento. Por no ofrecer información clara acerca de cómo obtienen las respuestas no son peores, de hecho, se asimilan más al razonamiento humano.

10. Razone y argumente a favor y en contra de usar modelos de redes CNN ya entrenados, y que se conocen han sido efectivos en otras tareas distintas de la que tiene que resolver, como modelos para aplicar directamente o como modelos a refinar para la tarea que tiene entre manos. Dar argumentos que no sean genéricos o triviales y que fundamenten su postura.

Dependiendo del problema que estemos tratando, lo relevante que sea obtener soluciones óptimas, el presupuesto o el tiempo para implementarlo, podríamos plantearnos elegir un sistema u otro.

Entrenar una red neuronal es difícil y requiere de tiempo y necesidad de cómputo. Es cierto que el resultado siempre será mejor, pero merece la pena probar si algún modelo ya entrenado resuelve nuestro problema ya que facilitaría todo el proceso enormemente.