

Examen de Computer Vision

Máster MIARFID, Universitat Politècnica de València, Mayo 2019

Apellidos:

Nombre:

30 minutos, sin apuntes

- 1 ☒ B Cuando se convierte de RGB a YCbCr se necesita calcular una suma ponderada de las componentes RGB. ¿Cuál de las siguientes es una conversión correcta para la componente Y, teniendo en cuenta las características visuales del sistema perceptivo humano?
- A) $Y = 0.299 R + 0.114 G + 0.587 B$
 - B) $Y = 0.299 R + 0.587 G + 0.114 B$
 - C) $Y = 0.587 R + 0.299 G + 0.114 B$
 - D) $Y = 0.587 R + 0.114 G + 0.299 B$
- 2 ☒ C Sean a y b dos niveles de gris, con $a < b$, los cuales definen un intervalo $[a, b]$ al cual le queremos aplicar ajuste de contraste (contrast stretching). Para ello se quiere obtener los nuevos valores de gris en el intervalo $[c, d]$. Los nuevos valores de gris para un punto p se calcularían como:
- A) $I'(p) = [I(p) - c] \frac{d-c}{b-a} + a$
 - B) $I'(p) = [I(p) - a] \frac{b-a}{d-c} + c$
 - C) $I'(p) = [I(p) - a] \frac{d-c}{b-a} + c$
 - D) $I'(p) = [I(p) - c] \frac{b-a}{d-c} + a$
- 3 ☒ A El detector de esquinas Harris está basado en los eigenvalores de la matriz de segundo orden calculada a partir de las derivadas de la imagen, suavizada con una ventana gaussiana. ¿Cuál es la condición distintiva de las esquinas?
- A) Los eigenvalores tienen que ser grandes y de similar magnitud
 - B) Los eigenvalores tienen que ser pequeños y de similar magnitud
 - C) Los eigenvalores tienen que ser de magnitudes diferentes: unos grandes y otros pequeños
 - D) Su producto tiene que ser inferior a $\alpha(\lambda_1 + \lambda_2)^2$, con $\alpha = 0,04$
- 4 ☒ C Dada una imagen de 32x32 píxeles de tres canales (RGB), se propone realizar la convolución (sin padding) con 16 filtros de 3x3. En cuanto a número de multiplicaciones, supondría realizar:
- A) 442.368
 - B) 43.200
 - C) 388.800
 - D) 27.648
- 5 ☒ D En el mismo caso anterior, el mapa de salida tendría las siguientes dimensiones:
- A) 16@32x32
 - B) 1@30x30
 - C) 1@32x32
 - D) 16@30x30

- 6 ☐ A En la fase de backpropagation en una red convolucional. Cuál de los siguientes operadores en general proporcionaría un Delta más disperso (con más ceros)?
- A) Max-pooling
 - B) Average-pooling
 - C) Convolución + activación lineal
 - D) Todos generarían la misma cantidad de ceros en el Delta retropropagado
- 7 ☐ D Con respecto a las DenseNet, la principal diferencia con la ResidualNet es:
- A) No tienen conexiones que “salten” hacia adelante
 - B) En general para la misma tasa de error tienen más parámetros
 - C) La versión stochastic-depth reduce considerablemente el número de parámetros
 - D) Los mapas de salida son concatenados en lugar de sumados
- 8 ☐ B Dada una imagen con un objeto cuyo ground-truth bounding-box va desde el pixel (1,1) hasta el pixel (100,100). Dado un detector que detecta dicho objeto pero con el bounding-box en (51,51)-(90,90). Cuál sería el valor del Intersection Over Union (IOU) ?
- A) IOU=0,08
 - B) IOU=0,16
 - C) IOU=0,138
 - D) IOU=0,32
- 9 ☐ B Con respecto a Region-CNN y al anterior estado del arte en detección (no basado en deep learning), podemos decir que R-CNN:
- A) Sustituye la parte de Region Proposal por una detector basado en CNN
 - B) Sustituye la parte de extractor de características por una CNN
 - C) Sustituye la parte del clasificador SVM por un CNN
 - D) Emplea Anchor Boxes para obtener mejor precisión en la localización
- 10 ☐ D Con respecto a Yolo. Siendo $S \times S$ el tamaño del grid en el que subdividimos la imagen, B el número de cajas (Boxes) a predecir en cada celda del grid, y C el número de clases. Yolo plantea el problema de detección como un regresor a un target-tensor de:
- A) $S \times S \times B \times C$
 - B) $S \times S \times (B + C)$
 - C) $S \times S \times 5 \times (B + C)$
 - D) $S \times S \times (B \times 5 + C)$

Examen de Computer Vision

Máster MIARFID, Universitat Politècnica de València, Recuperación Mayo 2019

Apellidos:

Nombre:

30 minutos, sin apuntes

- 1 ☐ B Dada una imagen en niveles de gris (sin color) de 32x32 píxels, se propone realizar la convolución (sin padding, stride=1) con 32 filtros de 5x5. En cuanto a número de multiplicaciones, supondría realizar:
- A) 19.600
 - B) 627.200
 - C) 25.088
 - D) 819.200
- 2 ☐ A En el mismo caso anterior, si empleáramos stride=2, el mapa de salida tendría las siguientes dimensiones:
- A) 32@14x14
 - B) 32@15x15
 - C) 32@13x13
 - D) 1@13x13
- 3 ☐ C En cuanto a la convolución traspuesta. Es una operación que
- A) Computacionalmente, el forward y el backward son iguales que para la convolución normal
 - B) Computacionalmente, sólo el forward es igual que el forward de la convolución normal
 - C) Computacionalmente, el forward es igual que el backward de la convolución normal
 - D) Computacionalmente, sólo el backward es igual que el backward de la convolución normal
- 4 ☐ A En una red bilineal, los mapas a los que se les aplica el producto externo son de 32@5x5 y 64@5x5. Por lo tanto el vector bilineal que se obtiene tiene el siguiente tamaño:
- A) 2.048
 - B) 25
 - C) 1.204
 - D) No se puede realizar este producto externo
- 5 ☐ B En la fase de backpropagation en una red convolucional. Cuál de los siguientes operadores en general proporcionaría un Delta más denso (con menos ceros)?
- A) Max-pooling
 - B) Average-pooling
 - C) ReLu
 - D) Todos generarían la misma cantidad de ceros en el Delta retropropagado

6 ☒ Con respecto a Region-CNN, Fast-RCNN aporta:

- A) Sustituye la parte de Region Proposal por una detector basado en CNN, la Region Proposal Network (RPN)
- B) Sustituye la parte de extractor de características por una CNN
- C) Realiza un único forward a la red y extrae las regiones de interés del mapa obtenido
- D) Emplea Anchor Boxes para obtener mejor precisión en la localización

7 ☒ Dada una imagen con un objeto cuyo ground-truth bounding-box va desde el pixel (11,11) hasta el pixel (90,90). Dado un detector que detecta dicho objeto pero con el bounding-box en (51,51)-(90,90). Cuál sería el valor del Intersection Over Union (IOU) ?

- A) IOU=0,25
- B) IOU=0,5
- C) IOU=0,125
- D) IOU=0,2

8 ☒ En esencia una U-Net es:

- A) Una red FCN (fully convolutional network) sin downsampling ni upsampling
- B) Una red convolucional con Fully Connected al final conectada a la función de pérdida
- C) Un GAN convolucional
- D) Una FCN (fully convolutional network) con topología encoder-decoder, downsampling y upsampling

9 ☒ Una diferencia importante de SSD con respecto a Yolo v1 es:

- A) En SSD una red neuronal previa es la que propone las regiones de interés. Es la Region Proposal Network
- B) En SSD se emplea un k-means sobre el groundtruth para definir los anchor boxes a emplear
- C) En SSD más de un mapa convolucional produce directamente un conjunto de detecciones. Por lo que están conectados directamente con la función de pérdida.
- D) En SSD se emplea una AlexNet en lugar de una VGG-16

10 ☒ En un problema de “image super-resolution” se suelen emplear topologías:

- A) FC, Fully Connected
- B) FCN, Fully Convolutional Networks
- C) Encoder (Convolutional) - Decoder (ConvolutionalT)
- D) DCGan, DeConvolutional GAN.