

VC – Convocatoria Ordinaria

Curso 2023-24

Indicaciones Generales

Esta prueba consiste en un examen tipo test con 20 preguntas. Tan solo se puede marcar una respuesta para cada una. Y dicha respuesta se marcará con una cruz en el formulario de respuestas (que se puede encontrar en la página 2).

Las preguntas correctas suman 1 punto. Las incorrectas descuentan 0.5 (es decir, dos preguntas mal descuentan una bien). La pregunta se puede dejar en blanco y ni sumará ni restará puntos. En cualquier caso, la calificación mínima que se puede obtener en esta prueba es 0.

La prueba tendrá una duración máxima de dos horas (finalizando a las 10:30).

Comentado [PM1]: Este examen se corresponde con el Tipo A. El examen Tipo B contenía exactamente las mismas preguntas, pero en distinto orden. El principal motivo de proceder de este modo fue el asegurarme de que ambas modalidades de examen tenían la misma dificultad.

Se indican en **negrita** las respuestas correctas para cada ejercicio/pregunta.



ugr

Universidad de Granada
Departamento de Ciencias de la Computación
e Inteligencia Artificial

Nombre y Apellidos:
DNI:

Formulario de respuestas

Utilice bolígrafo negro o azul para marcar la respuesta que considera correcta (a-d) para cada una de las preguntas (1-20) formuladas:



PREGUNTA	a	b	c	d
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				



ugr

Universidad de Granada
Departamento de Ciencias de la Computación
e Inteligencia Artificial

1) ¿Cuál diría que es el principal efecto que proporciona el siguiente filtro 2D?

$$\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$

- a) Suavizado Gaussiano
- b) Detección de Bordes
- c) **Interpolación Bilineal**
- d) Ninguna de las Anteriores

Comentado [PM2]: Extensión 2D del kernel de interpolación lineal comentado en slides 77-79 de https://pradograde2324.ugr.es/pluginfile.php/178475/mod_folder/content/0/2.3.Pyramids.pdf Un filtro de suavizado debería sumar 1 (es decir, estar normalizado).

2) Tenemos una red entrenada para un problema de clasificación, y queremos emplearla en un problema de regresión. ¿Cómo deberíamos hacer?

- a) Nada. Lo lógico es buscar una red pre-entrenada para regresión.
- b) Lo mejor que puedes hacer es convertir/adaptar tu problema de regresión en uno de clasificación, y reentrenar la red (o hacer finetuning) con este nuevo problema de clasificación.
- c) Cambia la última FC, pon otra FC adaptada a tu problema, y reentrena/finetunea usando cross-entropy loss y softmax a la salida.
- d) **Cambia la última FC, pon otra FC adaptada a tu problema, y reentrena/finetunea usando MSE y función de activación lineal a la salida.**

3) ¿Qué propiedad o teorema relativo a la convolución es el que, principalmente, permite la creación/composición de filtros más complejos a partir de filtros más sencillos?

- a) **Asociativa**
- b) Conmutativa
- c) Distributiva
- d) Teorema de la Convolución

Comentado [PM3]: Slide 35 de https://pradograde2324.ugr.es/pluginfile.php/178475/mod_folder/content/0/2.1.ImageFiltering.pdf

4) ¿Cuál sería la dimensionalidad del espacio de parámetros si quisiésemos detectar un objeto arbitrario con la Generalized Hough Transform incorporando también rotación y escalado uniforme?

- a) 2
- b) 3
- c) **4**
- d) 5

Comentado [PM4]: Slide 60 de https://pradograde2324.ugr.es/pluginfile.php/234073/mod_folder/content/0/3.2.TemplateMatching_Texture.pdf

5) ¿Qué tipo de filtro considera que es el siguiente?

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

- a) Suavizado Gaussiano
- b) Laplaciana de la Gaussiana
- c) **Sharpening**
- d) Ninguna de las anteriores

Comentado [PM5]: La LoG debería sumar cero, y en este filtro no ocurre. Véase slide 48-52 de https://pradograde2324.ugr.es/pluginfile.php/178475/mod_folder/content/0/2.1.ImageFiltering.pdf para información sobre Sharpening



ugr

Universidad de Granada
Departamento de Ciencias de la Computación
e Inteligencia Artificial

- 6) A continuación, se muestra lo que podría ser el pseudocódigo de una convolución separable. I es la imagen a filtrar (con dimensiones $width \times height$), $*$ la multiplicación, w es la longitud (impar) del kernel, $\tilde{w} = \frac{w-1}{2}$, y g_h y g_v serían los filtros 1D horizontal y vertical, respectivamente. Asumiendo que I ya contiene el padding necesario, es decir, que la convolución se realiza en modo 'valid'. ¿Es dicho pseudocódigo correcto?

```
CONVOLVESEPARABLE( $I, g_h, g_v$ )
  for  $y \leftarrow 0$  to  $height - 1$  do
    for  $x \leftarrow \tilde{w}$  to  $width - 1 - \tilde{w}$  do
      val  $\leftarrow 0$ 
      for  $i \leftarrow 0$  to  $w - 1$  do
        val  $\leftarrow val + g_h[i] * I(x + \tilde{w} - i, y)$ 
      tmp( $x, y$ )  $\leftarrow val$ 

      for  $y \leftarrow \tilde{w}$  to  $height - 1 - \tilde{w}$  do
        for  $x \leftarrow 0$  to  $width - 1$  do
          val  $\leftarrow 0$ 
          for  $i \leftarrow 0$  to  $w - 1$  do
            val  $\leftarrow val + g_v[i] * tmp(x, y + \tilde{w} - i)$ 
          out( $x, y$ )  $\leftarrow val$ 

  return out
```

- El pseudocódigo es correcto y realiza, adecuadamente, la convolución separable.
- El pseudocódigo es incorrecto porque, cuando pasamos el segundo filtro, no hemos guardado los resultados del primero.
- El pseudocódigo es incorrecto porque no desplazamos adecuadamente los filtros por columnas y por filas.
- El pseudocódigo es incorrecto, pero por motivos distintos a los listados más arriba.

Comentado [PM6]: Véase materiales de Gaussian Discretization de Stan Birchfield.
https://pradograde2324.ugr.es/pluginfile.php/178298/mod_folder/content/0/Material%20de%20Apoyo/Gaussian_discretization.pdf
 (pág. 4)

- 7) ¿Cuál de las siguientes afirmaciones es correcta?

- Una homografía es una matriz 3x3 con 8 grados de libertad
- Una homografía es una matriz cuyo determinante es distinto de cero
- Una homografía relaciona dos imágenes de la misma superficie
- Todas las anteriores son correctas

- 8) ¿Cuál es el principal objetivo de los *connecting paths* empleados en las redes U-Net para segmentación de imágenes?

- Combinar información espacial (presente en las últimas capas de la red) con información semántica (presente en las primeras capas de la red).
- Combinar información espacial (presente en las primeras capas de la red) con información semántica (presente en las últimas capas de la red).
- Al igual que en ResNet, aprender el residual mapping, facilitando el aprendizaje de la identidad.
- Aliviar el vanishing gradient problem.

Comentado [PM7]: Slide 20 de
https://pradograde2324.ugr.es/pluginfile.php/234102/mod_folder/content/0/5.1.Image_Segmentation.pdf



9) ¿Cuántos parámetros/pesos, incluyendo bias, tiene la siguiente red neuronal?

Layer Type	Kernel Size (for conv layers)	Input Output dimension	Input Output channels (for conv layers)
Conv	11x11	224x224 214x214	3 68
BN	-	214x214 214x214	-
ReLU	-	214x214 214x214	-
MaxPooling	2x2	214x214 107x107	-
Conv	7x7	107x107 101x101	68 10
BN	-	101x101 101x101	-
ReLU	-	101x101 101x101	-
Dropout	-	101x101 101x101	-
FC	-	102010 50	-
ReLU	-	50 50	-
FC	-	50 10	-

- a) 6977686
- b) 5159160
- c) 5159454
- d) Ninguno de los anteriores

Comentado [PM8]: $(11 \cdot 11 \cdot 3 + 1) \cdot 68 + 2 \cdot 68 + (7 \cdot 7 \cdot 68 + 1) \cdot 10 + 2 \cdot 10 + 102010 \cdot 50 + 50 + 50 \cdot 10 + 10 = 5159298$

10) Dentro del contexto de template matching, ¿puede la suma de diferencias al cuadrado (SSD), entre un template H y el patch F de una imagen, implementarse con filtrado lineal?

- a) No. Es una operación que no puede traducirse a filtros lineales.
- b) Sí, podríamos calcularla como una constante (H^2) menos 2 veces la correlación de H y F, más la correlación de F^2 y un kernel de unos.
- c) Sí, podríamos calcularla como una constante (H^2) más 2 veces la correlación de H y F, menos la correlación de F^2 y un kernel de unos.
- d) Sí, podríamos sencillamente minimizar la correlación de H y F.

Comentado [PM9]: Slide 29 de https://pradogrado2324.ugr.es/pluginfile.php/234073/mod_folder/content/0/3.2.TemplateMatching_Texture.pdf



ugr

Universidad de Granada
Departamento de Ciencias de la Computación
e Inteligencia Artificial

11) ¿Cuál es, posiblemente, la principal limitación de la aproximación Bag-of-Words (BoW) en visión por computador?

- a) Complejidad computacional elevada.
- b) Dificultad para crear el vocabulario visual con técnicas de aprendizaje supervisado.
- c) Sensibilidad a variaciones en iluminación, posición y orientación de los objetos de la imagen.
- d) **Los histogramas de palabras visuales suprimen toda información espacial, dificultando p.ej. la localización de objetos en imágenes.**

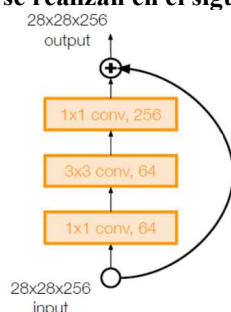
Comentado [PM10]: Slide 43 en https://pradogrado2324.ugr.es/pluginfile.php/234073/mod_folder/content/0/3.3.BoW_Blobs.pdf

12) Imaginemos una pirámide Gaussiana en cuya base no se encuentra la imagen original (es decir, el nivel-0 no incluye la imagen original). ¿Es posible reconstruir dicha imagen a partir exclusivamente de la pirámide Gaussiana?

- a) Sí, podemos ir expandiendo, haciendo upsampling de cada uno de los niveles, hasta obtener la imagen original.
- b) **No, no es posible.**
- c) Sí, pero depende del sigma que empleemos para suavizar.
- d) Sí, pero depende del tipo de interpolación empleada a la hora de submuestrear.

Comentado [PM11]: Necesitaríamos una pirámide Laplaciana (construida a partir de una pirámide Gaussiana en la que la imagen original estuviese en la base).

13) ¿Cuántas multiplicaciones se realizan en el siguiente bloque residual?



- a) **~54M**
- b) ~231M
- c) ~42M
- d) Ninguno de los anteriores

Comentado [PM12]: SIN BIAS:
 $(1 \cdot 1 \cdot 256) \cdot 28 \cdot 28 \cdot 64 + (3 \cdot 3 \cdot 64) \cdot 28 \cdot 28 \cdot 64 + (1 \cdot 1 \cdot 64) \cdot 28 \cdot 28 \cdot 256 = 54\,591\,488$ multiplicaciones

CON BIAS (aunque no tiene demasiado sentido, porque la entrada al bias es la constante '1', de modo que, como tal, no sería necesaria ninguna multiplicación):
 $(1 \cdot 1 \cdot 256 + 1) \cdot 28 \cdot 28 \cdot 64 + (3 \cdot 3 \cdot 64 + 1) \cdot 28 \cdot 28 \cdot 64 + (1 \cdot 1 \cdot 64 + 1) \cdot 28 \cdot 28 \cdot 256 = 54\,892\,544$ multiplicaciones

14) Imaginemos que disponemos de un volumen de entrada de tamaño 28x28x128. Si ahora realizamos la convolución con 96 filtros 5x5, con padding=1 y stride=5. ¿Cuáles serán las dimensiones del volumen de salida?

- a) 96x5x5
- b) **96x6x6**
- c) 128x5x5
- d) 128x6x6

Comentado [PM13]: $\text{Output_size} = ((N+2 \cdot P-F)/\text{stride}) + 1$

En nuestro caso, esto se traduciría en

$$(28+2 \cdot 1-5)/5 + 1 = 25/5 + 1 = 5 + 1 = 6$$

Por lo que el volumen de salida sería de 96 filtros de 6x6.

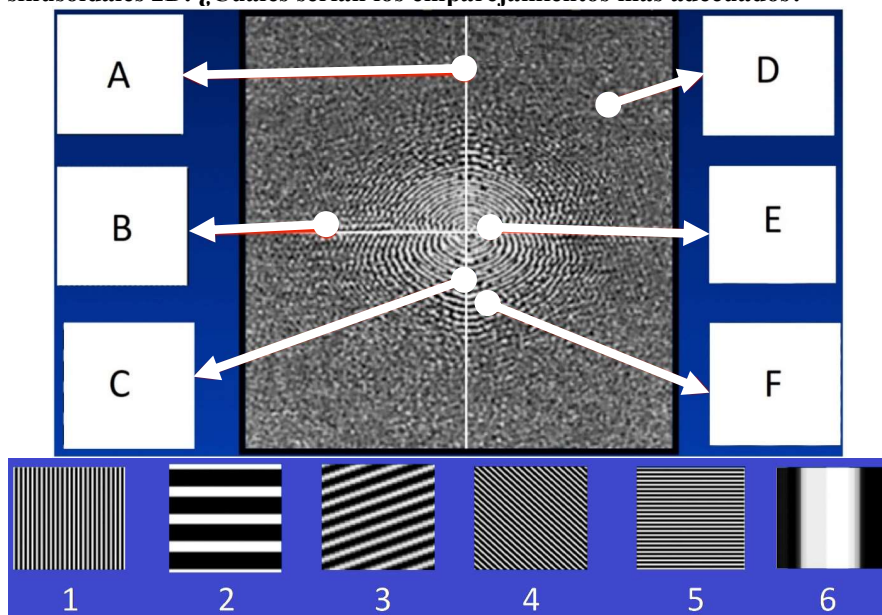
Si preferimos emplear una representación visual e intuitiva, veríamos que vamos desplazando un kernel 5x5 sobre la imagen de 30x30 (una vez incorporado el padding), de forma que el kernel va ocupando posiciones disjuntas y generando la imagen de 6x6 a la salida.



ugr

Universidad de Granada
 Departamento de Ciencias de la Computación
 e Inteligencia Artificial

- 15) Consideremos la siguiente imagen de Fourier, así como las siguientes 6 “intensity images” representando distintas frecuencias y patrones sinusoidales 2D. ¿Cuáles serían los emparejamientos más adecuados?



- a) A-2, B-6, C-5, D-3, E-1, F-4
b) **A-5, B-1, C-2, D-4, E-6, F-3**
c) A-1, B-5, C-6, D-4, E-2, F-3
d) A-6, B-2, C-1, D-3, E-5, F-4

Comentado [PM14]: Slide 16 de https://pradogrado2324.ugr.es/pluginfile.php/178475/mod_folder/content/0/2.3.Pyramids.pdf

- 16) ¿Cuál de las siguientes afirmaciones sobre las propiedades del descriptor SIFT es correcta?

- a) Es invariante ante cambios del punto de vista de adquisición de la fotografía
b) **Es invariante a rotaciones y cambios de escala**
c) Es invariante ante oclusiones y cambios de intensidad
d) Todas las anteriores son correctas

Comentado [PM15]: Es robusto a cambios de punto de vista, pero a partir de 30° los resultados pueden ser bastante malos. Véase slides 47-49 de https://pradogrado2324.ugr.es/pluginfile.php/234073/mod_folder/content/0/3.4.Harris_SIFT_HOG.pdf

Comentado [PM16]: En todo caso, sería robusto y no invariante.

- 17) Si suavizamos una imagen tres veces consecutivas con un filtro Gaussiano σ , ¿cuál sería el sigma equivalente que nos permitiría obtener el mismo resultado suavizando una única vez?

- a) $\sigma\sqrt{3}$
b) $\sqrt{3}\sigma$
c) $3\sqrt{\sigma}$
d) Ninguna de las anteriores



18) ¿Cuál de las siguientes afirmaciones sobre los algoritmos de Canny y Otsu es correcta?

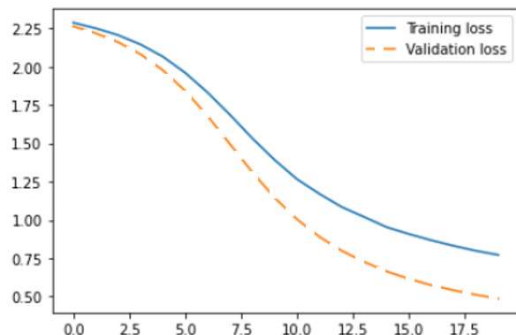
- a) El detector de bordes de Canny se comporta como un filtro lineal dado que emplea filtrado Gaussiano para suavizar/emborronar la imagen y, a continuación, emplea filtrado lineal para calcular el gradiente.
- b) El umbral óptimo en el algoritmo de Otsu se obtiene minimizando la varianza inter-clase o maximizando la varianza intra-clase.
- c) En el algoritmo de Canny obtendremos bordes más discontinuos/fragmentados si reducimos el umbral bajo.
- d) Las tres respuestas anteriores son erróneas.

Comentado [PM17]: Falso. Aunque Canny hace todas estas cosas, también incorpora operaciones no lineales, como thresholding, hysteresis, non-maximum suppression.

Comentado [PM18]: Es al revés: maximizar varianza inter-clase o minimizar varianza intra-clase.

Comentado [PM19]: Realmente sería al revés. Si reducimos el umbral bajo, habría más bordes que aceptaríamos/extenderíamos en el proceso de hysteresis. Obtendríamos bordes más discontinuos si lo subiésemos, dado que estaríamos siendo más exigentes con los potenciales bordes a aceptar.

19) Imaginemos que tenemos las curvas de aprendizaje que se muestran a continuación. ¿Cuál de las siguientes interpretaciones considera más pertinente y certera?



- a) Estamos presenciando un fenómeno de sobreentrenamiento, en donde las curvas de entrenamiento y validación, cada vez, se van separando más.
- b) El conjunto de validación posiblemente es más “fácil” que el de training.
- c) El learning rate seguramente sea demasiado bajo, motivo por el cual estamos avanzando muy lentamente en el proceso de optimización.
- d) El inicio del entrenamiento parece excesivamente lento (probablemente motivado por una inicialización de pesos inadecuada).

20) ¿Cuál de las siguientes afirmaciones es correcta?

- a) Harris usa primeras derivadas en la second moment matrix (SMM) para detectar esquinas, mientras que SIFT usa segundas derivadas para detectar regiones.
- b) SIFT emplea una aproximación multiescala en un espacio de escalas Gaussiano, y propone un descriptor basado en histogramas locales construidos a partir de la orientación del gradiente.
- c) Harris propone un enfoque más fotométrico que geométrico, por lo que, aunque haya pocos objetos en la escena, podríamos obtener suficientes esquinas como puntos de interés.
- d) Todas las anteriores son correctas.

Comentado [PM20]: Recuérdese que SIFT detecta regiones usando la diferencia de Gaussianas, que es una aproximación a la Laplaciana de la Gaussiana, y la LoG emplea las segundas derivadas. Véase slides 55-63 de https://pradograde2324.ugr.es/pluginfile.php/178475/mod_folder/content/0/2.2.GaussianFilters_Edges.pdf y 39-41 en https://pradograde2324.ugr.es/pluginfile.php/234073/mod_folder/content/0/3.4.Harris_SIFT_HOG.pdf

Comentado [PM21]: Las esquinas son, principalmente, efectos fotométricos derivados de la iluminación. Puedes no tener objetos geométricos bien definidos y, a pesar de ello, detectar numerosas esquinas (véase, por ejemplo, la imagen de una determinada textura).



ugr

Universidad de Granada
Departamento de Ciencias de la Computación
e Inteligencia Artificial