VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

Datos pequeños, modelos grandes, y la recompensa del razonamiento

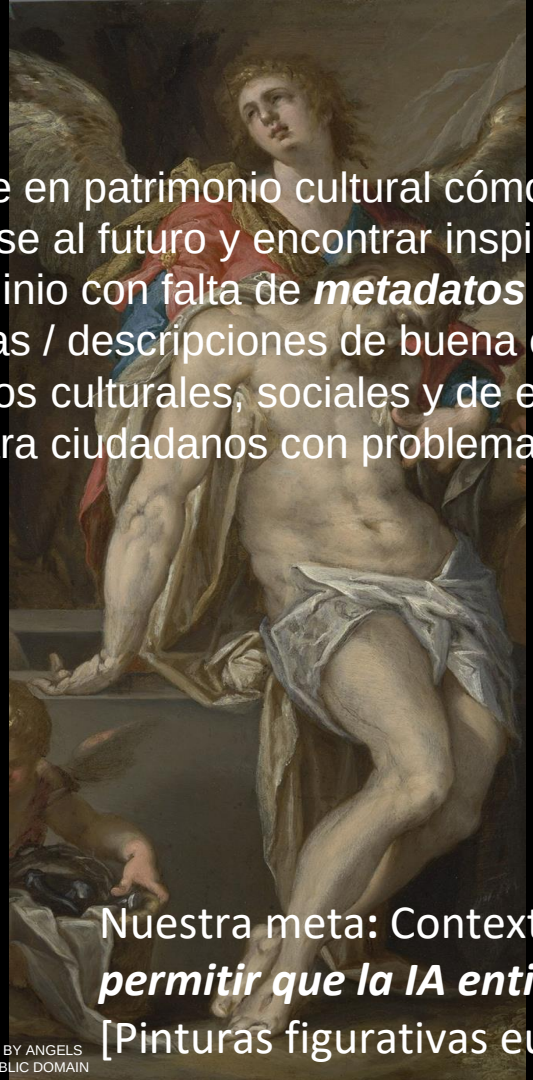
Maria-Cristina Marinescu

IQS
UNIVERSITAT RAMON LLULL

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Saint George on a Bike: Motivación

- Enfoque en patrimonio cultural cómo manera de entender nuestro pasado, acercarse al futuro y encontrar inspiración.
- Un dominio con falta de ***metadatos*** (de calidad)!
- Etiquetas / descripciones de buena calidad fomentan la investigación, proyectos culturales, sociales y de educación, y mejoran la accesibilidad Web para ciudadanos con problemas de vista.



Anotación automatica (metadatos)

Nuevas formas de interaccionar con usuarios via paginas Web o aplicaciones

Interacción con minorías, e.g. ciudadanos con discapacidad visual

Mejorar las busquedas y la navegación

VIII JORNADA DE PATRIMONI CULTURAL

Nuestra meta: Contextualizar los objectos y la composición imagistica para ***permitir que la IA entienda la cultura, los símbolos y la tradición***.
[Pinturas figurativas europeas, muchas iconográficas, XII-XVIII]

THE BODY OF CHRIST SUPPORTED BY ANGELS BARTHOLOMAEUS SPRANEERS. PUBLIC DOMAIN

¿Porqué no usar las herramientas existentes? (recuerda, esto era 2018!!)

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

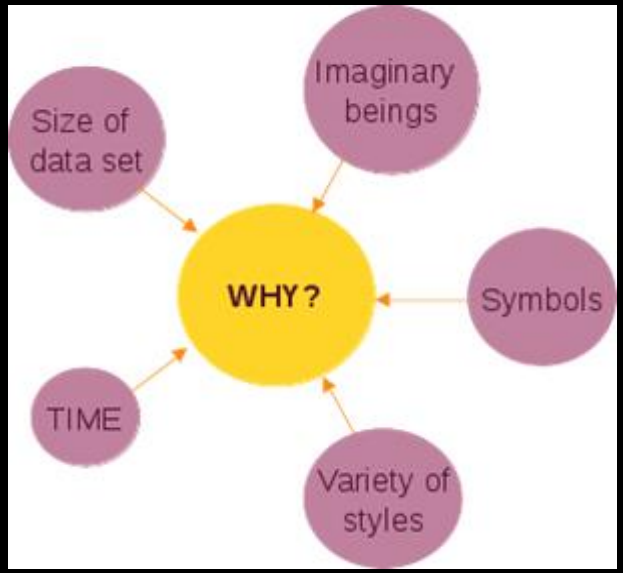
Desafío principal:
Soluciones actuales funcionaban muy bien para imágenes de día-a-día, pero no para imágenes de patrimonio cultural (PC). El motivo es que estuvieron entrenados en bases de datos muy grandes, con imágenes actuales.

¿Concretamente, de donde surge el problema?

E.g.

- Objetos antiguos que ya no se usan, ej. tintero
- Objectos que tenían otra forma en el pasado, ej. arado
- Nuevos objectos, distintos, pero con la misma forma que objetos antiguos, ej. teléfono móvil/libro
- Acciones inusuales (que no se retratan en fotografías), ej. un hombre matando un caballo

Aplicar tecnicas de distintos campos (de IA) a imágenes o (imagen, texto): deep learning, modelos basados en el procesamiento de lenguaje natural, extracción de metadatos semánticos y razonamiento



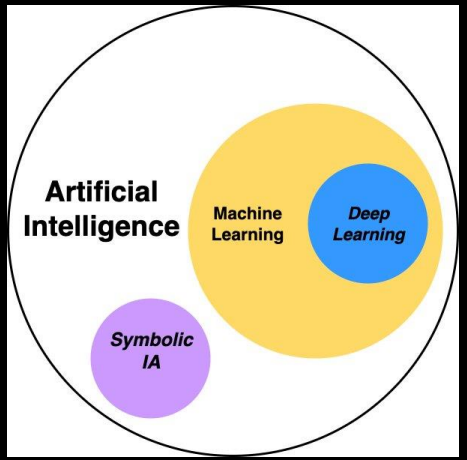
Objetos imaginarios
Símbolos
Distintos estilos
Objetos anacrónicos
Número reducido de datos

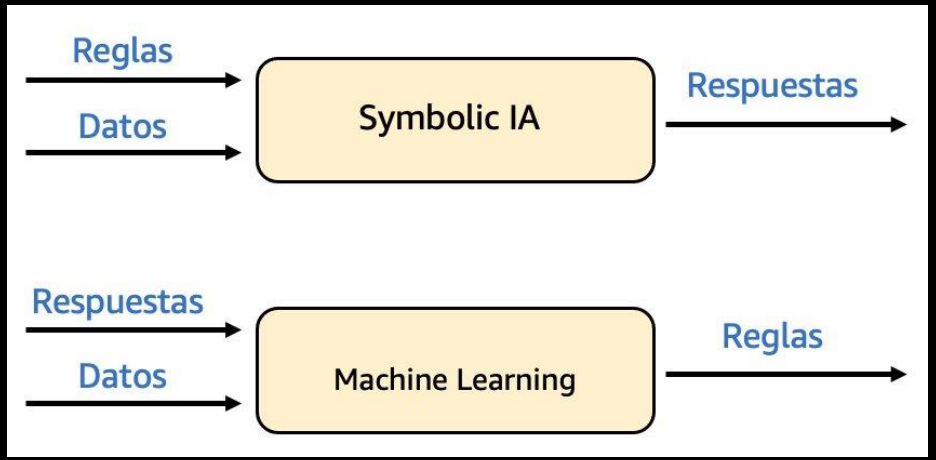
¿Concretamente, de donde surge el problema?

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

- Deep learning
- Procesamiento de lenguaje natural
- Tecnologías semánticas (y razonamiento)





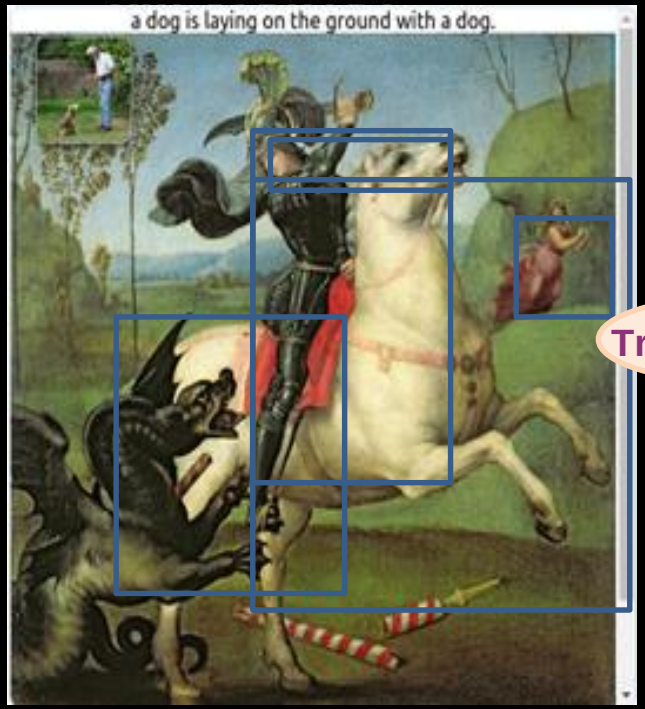
Datos de entrada y salida

Entrada: imagen, posiblemente metadatos

Salida: distintos niveles semánticos

Semantic Level	Examples
Semantic resources (tags) From vocabularies, preferably with linked data URIs. Anotación de clases	Adoration of the Magi: • Jesus Christ, Virgin Mary, Wise Man (as subjects coming from a vocabulary). • http://iconclass.org/rkd/73B57/: "Adoration of the kings: the Wise Men present their gifts to the Christ-child (gold, frankincense and myrrh)."
Textual captions **Generación de descripciones**	"Man reading a book in a dark room." "Woman plays a guitar outdoors during sunny weather."
Semantic/knowledge graph Graphs with relationships between semantic resources, where the link can also have a URI.	(St. George, kill, dragon) (Woman, sits) Adoration of the Magi: (Wise Man, adore, Jesus Christ), (Virgin Mary, hold, Jesus Christ) Triplas (s p o)

Dos maneras de generar anotaciones ricas y de calidad



San Jorge, cabalgando, mata el dragón. En el fondo, la princesa está corriendo.

Triplas (s,p,o)

«Semillas» para descripciones simples:
objectos: caballero, espada, caballo, dragón, mujer
semillas: (caballero mata dragón), (caballo cabalga caballo), (mujer corre)
Porqué este acercamiento: no existen (muchas, apropiadas) descripciones del contenido visual de imágenes de PC
Qué implica esta alternativa: 1. obtener anotaciones de objetos para el entrenamiento 2. generar relaciones probables entre objetos

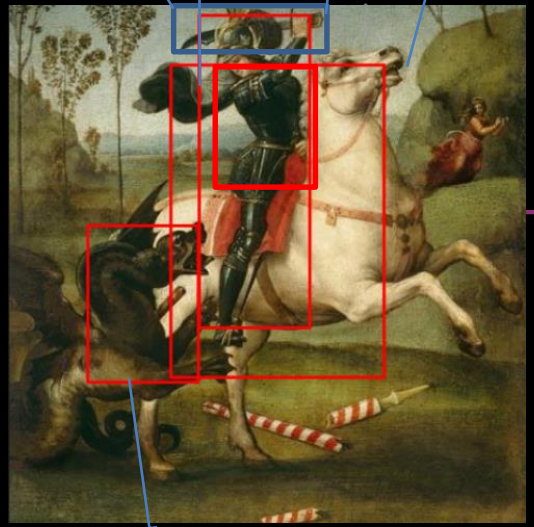
Anotación de clases

Descripciones del contenido visual, en lenguaje natural:
Que implica esta alternativa: obtener anotaciones de imágenes en forma de descripciones en lenguaje natural, para el entrenamiento

Generación de descripciones

Detección de objetos con deep learning

Refinamiento de objectos y relaciones via razonamiento semantico o PLN



espada armadura persona caballo

perro

Persona

Persona con espada es ___

Caballero / rey

Caballero lleva espada

Espada

Caballero ____ espada

Generación de relaciones

Armadura

Caballero/rey con armadura es ___

Caballero

Refinamiento de clases

Refinamiento de relaciones

Caballo

Caballero en caballo

Caballero cabalga caballo

Perro

Caballero ____ perro

???

Para personas: puede generar descripciones simples

Para computadoras: puede utilizar triplas para formar estructuras conectadas (grafos de conocimiento)

Datos para la generación de descripciones visuales

- Se pueden generar triplas (sujeto, relación, objeto) o acciones ej. sentado, comer etc.
- Para generar descripciones complejas en lenguaje natural se necesita un conjunto de datos **suficientemente grande de (descripción, imagen)** para entrenar un modelo de deep learning

Utilizar un clasificador para seleccionar frases que explican el contenido visual de imágenes, desde, ej. El Prado

(1) problemas con el lenguaje complejo, difícil para PLN (Procesamiento de Lenguaje Natural)
(2) se asume que VEMOS y no hay necesidad de explicar el contenido visual

Plataforma Zooniverse para crowdsourcing

7543 imágenes anotadas con 4-5 descripciones

Nuestra meta es la anotación de todas 15K imágenes con 5 anotaciones por image

Directrices de anotación desarrolladas e implementadas

1,859 voluntarios

154 hilos de discusion

362 comentarios

17 mencioens en media y web

Generación de descripciones complejas (lenguaje natural)

Entrenamiento usa mecanismo llamado de «atención»: la idea es **correlacionar áreas específicas de la imagen con palabras de la descripción.**

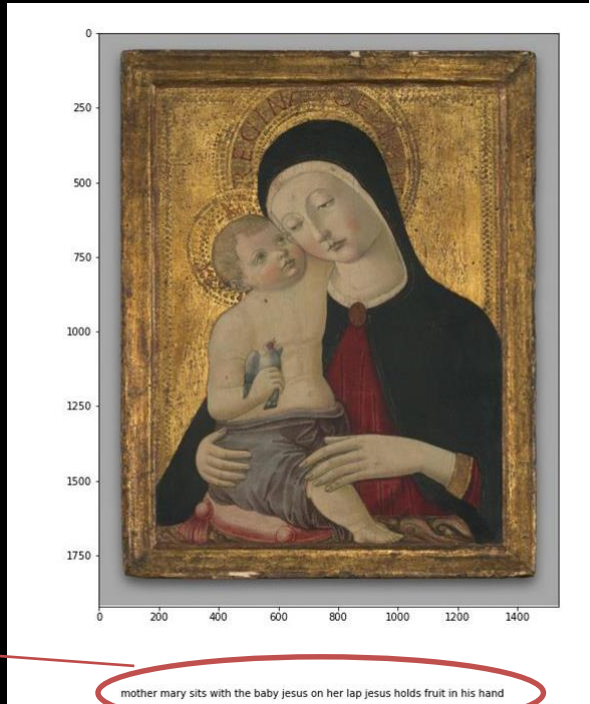
Con nuestras anotaciones para 7500 imágenes, generamos descripciones buenas para imágenes no muy complejas (ej. retratos, escenas bíblicas con pocos detalles). Para escenas complejas, se necesitan mas anotaciones!

Halflength elderly man in fur coat and jacket looks at the viewer.

Hombre anciano de medio cuerpo en abrigo de piel y chaqueta mira al espectador.

Mother Mary sits with the baby Jesus on her lap. Jesus holds fruit in his hand.

La Madre María se sienta con el niño Jesús en su regazo. Jesús sostiene la fruta en su mano.



VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Desafíos

- Colección de datos
- Metadatos de pobre calidad o inexistentes
- Métodos de evaluación

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

Colección de datos

Ej. (para imágenes):

- Algunas clases tienen poca representación
- Estilo, medio, color muy distintos
- Numero de pinturas no suficiente (estándares IA) y no se pueden producir más cuando es necesario

Como resolver los desafíos:

- Conjunto pequeño de datos requiere el uso de técnicas complementarias, particularmente para detectar objetos inusuales / imaginarios / simbólicos
- Técnicas de augmentación de datos no funcionan muy bien para PR

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

Metadatos inexistentes o de pobre calidad

Ej. (texto):

- No hay muchas descripciones del contenido visual, tampoco anotaciones exhaustivas de objetos.
- Las descripciones existentes contienen mayormente información de contexto o forma, y mucho menos información sobre el contenido – se asume que la imagen SE VE y ya se sabe de qué se trata.
- No existe información «formal» sobre cuáles son las relaciones de tipo visual (ej. cabalga, vuela, esta sentada)- ej. una ontología.

¿Como tratar estas deficiencias?:

- Clasificar si una frase trata sobre contenido visual o no.
- Aproximar relaciones visuales en PC con las que aparecen en recursos como anotaciones de fotografías (e.g. COCO) o *IconClass*
- Crowdsourcing

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

Metodos de evaluación

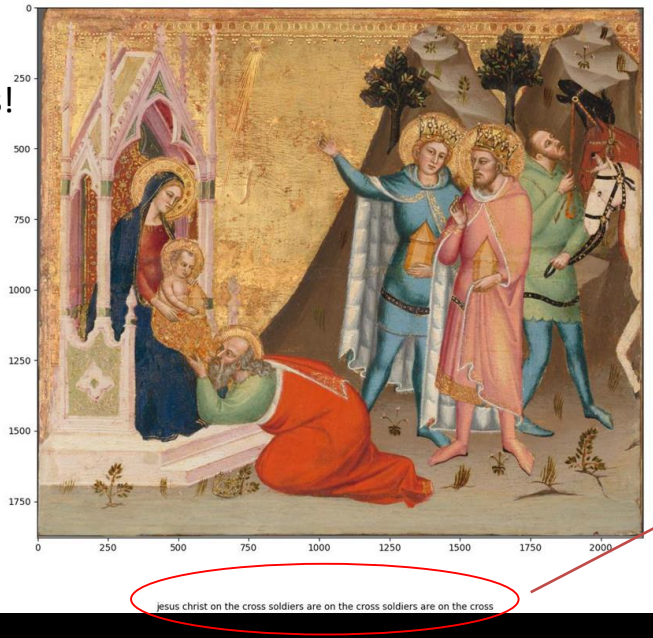
Metadatos generados automáticamente vs. por humanos

- La evaluación automática con métricas existentes es problemática para descripciones, especialmente dada la diversidad que pueda haber para un contenido lleno de símbolos, santos o interpretaciones debidas al conocimiento más o menos profundo de la iconografía y el arte
- Importa tanto cuantificar la calidad de los metadatos, como la utilidad para el usuario

La pregunta es: son las técnicas «bottom up» (ej. deep learning) suficientes para generar descripciones de PC de alta calidad?

¿Son las técnicas «bottom up» suficientes para generar descripciones de PC de alta calidad?

Halucinaciones!



Adoración de los magos, no la crucificación.

Jesus Christ on the cross. Soldiers are on the cross.

Muy probablemente se necesita una mezcla de bottom-up (deep learning) y top-down (semántica: grafos de conocimiento + razonamiento, PLN) para modelar correctamente el conocimiento en el dominio de PC

Ejemplo: FrAI Angelico (liderado por Quim Moré)

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

Prueba de concepto con pinturas de El Prado, usa mas información que SGoaB:

- Tuplas (sujeto, predicado, objeto) extraídas de las anotaciones de **Iconclass** y descripciones de pinturas de las colecciones de El Prado
- Etiquetas extraídas de los XMLs de los metadatos asociados a imágenes de El Prado

```
Attribute indentifiers for: ('person_1', 'is_with')
halo_1
```

	subj	pred	dobj	pobj	topic (with Iconclass code)
279	Christ	is_with	halo		Christ(11D)

	subj	pred	dobj	pobj	topic (with Iconclass code)
818	the_virgin_Mary	is_with	halo		The_Virgin_Mary(11F)

Algunas conclusiones (SGoaB)

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

El proyecto descubre nuevos desafíos en PC, que llevan a nuevas preguntas de investigación.

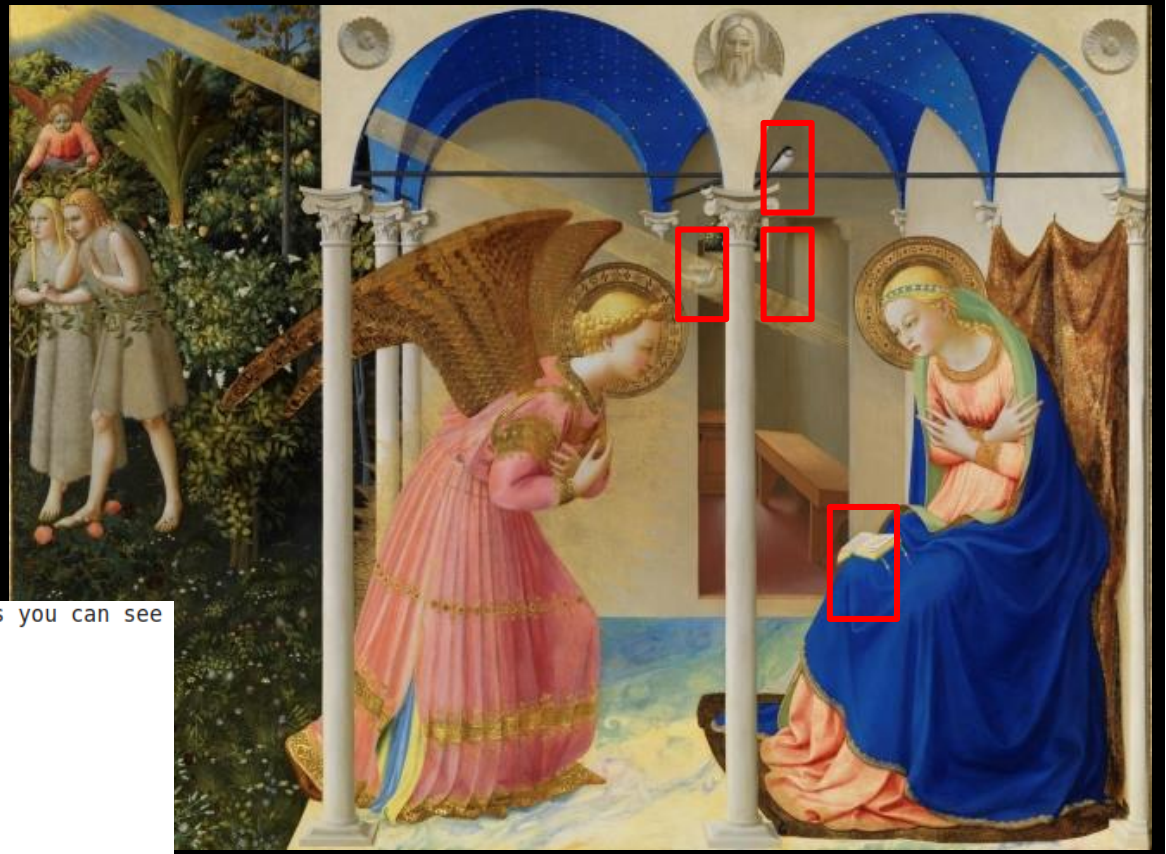
- Ej. Dado el tamaño del conjunto de datos en PC y la cantidad de descripciones manuales que se pueden utilizar para entrenamiento, como y hasta que punto se pueden complementar los metodos bottom-up con conocimiento top-down para ***prevenir las alucinaciones***?

- Ej. Son las descripciones simples - vía triplas - redundantes, si existen descripciones en lenguaje natural?

- Ej. Que pasa con los ciudadanos con discapacidad visual si las instituciones artísticas (ej. museos) proporcionan descripciones asumiendo visión? Bajo este presupuesto, como pueden las computadoras aprender a «ver»? SGoaB ayuda con la ***inclusión de la ciudadanía en el ámbito del patrimonio cultural.***

Descubrimiento de entidades no reconocidas

Se pide al visitante de la web que mire detalladamente el cuadro y marque las entidades que ve y que no han sido identificadas



Look at the painting more closely and tick the entities you can see

- [x] paloma_blanca_simbólica
- [ ] azucena
- [x] libro
- [x] golondrina
- [x] vidriera_artística

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

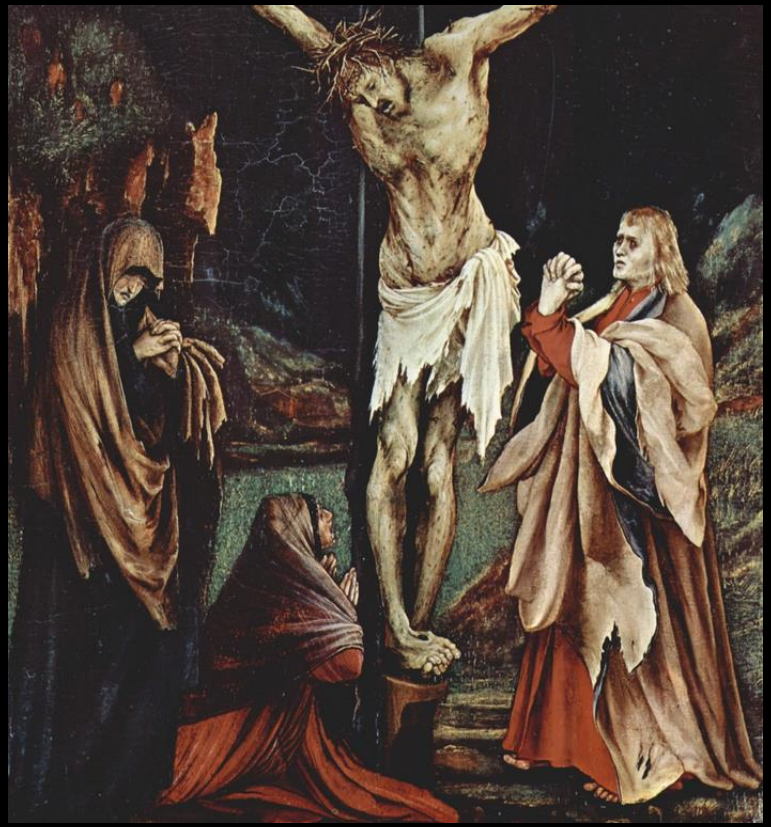
Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

¿Y qué hay de los recientes large language models (LLMs, e.g. GPT4)?

Podríamos directamente usar estas nuevas herramientas y pedirles que describan el contenido de una imagen?

Las descripciones de imágenes típicas (no sorprendentes) son muy buenas (cosa que no ocurría hace 1.5 años), pero todavía producen alucinaciones.

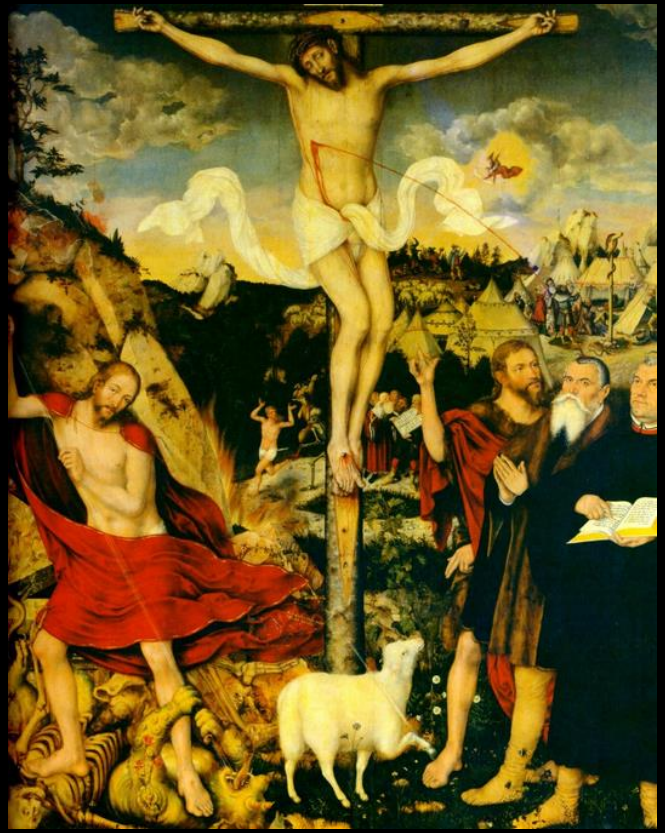
¿Y qué hay de los recientes large language models (LLMs, e.g. GPT4)?



"… Las figuras se muestran con halos, indicando su posicioón sacra."

Descripción de objetos inexistentes, hipotéticamente porque normalmente están presentes en el tópico / tema que se reconoce.

¿Y qué hay de los recientes large language models (LLMs, e.g. GPT4)?



"A la izquierda, una figura está arrodillada con un pecho expuesto, quien es San Juan el Evangelista, a menudo ilustrado de una manera joven y llena de compasión. A la derecha hay dos figuras …"

Describe cosas imaginarias – arrodillado
No describe objetos que sí existen, hipotéticamente porque normalmente están ausentes en el tópico / tema reconocido – matanza de un animal, de pie pisando un esqueleto.

VIII JORNADA DE PATRIMONI CULTURAL

LA INTEL·LIGÈNCIA ARTIFICIAL APLICADA AL PATRIMONI CULTURAL

Generalitat de Catalunya
Agència Catalana del Patrimoni Cultural

¿Y qué hay de los recientes large language models (LLMs, e.g. GPT4)?

Podríamos directamente usar estas nuevas herramientas y pedirles que describan el contenido de una imagen?

Las descripciones de imágenes típicas (no sorprendentes) son muy buenas (cosa que no ocurría hace 1.5 años), pero todavía producen alucinaciones.

Podríamos mejorar la descripción si le pasamos al LLM prompt basados en los objetos / triplas que extraemos?
... trabajo en curso

¿Preguntas?

mariacristina.marinescu@gmail.com