

Competencia Kaggle

Detección SIIM-FISABIO-RSNA COVID-19

Identificar y Localizar anomalías de COVID-19 en radiografías de tórax



Objetivo General:

Conformar la base de un equipo para proyectos de AI del Centro de Graduados para proyectos generales

Proyecto KAGGLE:

General Identificar y Localizar anomalías de COVID-19 en radiografías de tórax.

1-Equipo Final:

Guillermina, Gustavo, Franklin, David, Cristian Hugo Jaca

Sin respuesta

54 9 11 6401 7671 Rosana Mac Lean
57 318 8337153 Jennyfer Combariza
54 9 3516 640100 Walter Nieto
54 9 3512 44 3442 Cra Sol Figueroa

2-El proyecto fuera de la competencia

Luego del desarrollo el pedido de Omar : es una presentación. Hacemos evento por zoom y seguimos avanzando para presentación potenciales clientes.

3-El proyecto dentro de la competencia



Cinco veces más mortal que la gripe, COVID-19 causa una morbilidad y mortalidad significativas. Al igual que otras neumonías, la infección pulmonar con COVID-19 produce inflamación y líquido en los pulmones. COVID-19 se parece mucho a otras neumonías virales y bacterianas en las radiografías de tórax, lo que dificulta su diagnóstico. Su modelo de visión por computadora para detectar y localizar COVID-19 ayudaría a los médicos a brindar un diagnóstico rápido y confiable. Como resultado, los pacientes podrían recibir el tratamiento adecuado antes de que se produzcan los efectos más graves del virus.

Actualmente, COVID-19 se puede diagnosticar mediante la reacción en cadena de la polimerasa para detectar material genético del virus o una radiografía de tórax. Sin embargo, pueden pasar algunas horas y, a veces, días antes de que se obtengan los resultados de las pruebas moleculares. Por el contrario, las radiografías de tórax se pueden obtener en minutos. Si bien existen pautas para ayudar a los radiólogos a diferenciar el COVID-19 de otros tipos de infección, sus evaluaciones varían. Además, los no radiólogos podrían recibir apoyo con una mejor localización de la enfermedad, como con un cuadro delimitador visual.

Como organización líder en el cuidado de la salud en su campo, la misión de la Sociedad de Informática por Imágenes en Medicina (SIIM) es promover la informática de imágenes médicas a través de la educación, la investigación y la innovación. SIIM se ha asociado con la Fundación para el Fomento de la Salud y la Investigación Biomédica de la Comunitat Valenciana (FISABIO), el Banco de Datos de Imágenes Médicas de la Comunitat Valenciana (BIMCV) y la Sociedad Radiológica de Norteamérica (RSNA) para este concurso.

En esta competencia, identificará y localizará anomalías de COVID-19 en radiografías de tórax. En particular, categorizará las radiografías como negativas para neumonía o típicas, indeterminadas o atípicas para COVID-19. Nuestro grupo y nuestro modelo trabajarán con datos de imágenes y anotaciones de un grupo de radiólogos.

Esto, ayudará a los radiólogos a diagnosticar a los millones de pacientes con COVID-19 con mayor confianza y rapidez. Esto también permitirá a los médicos ver el alcance de la enfermedad y les ayudará a tomar decisiones con respecto al tratamiento. Dependiendo de la gravedad, los pacientes afectados pueden necesitar hospitalización, ingreso en una unidad de cuidados intensivos o terapias de apoyo como ventilación mecánica. Como resultado de un mejor diagnóstico, más pacientes recibirán rápidamente la mejor atención para su afección, lo que podría mitigar los efectos más graves del virus.

link al proyecto

https://www.kaggle.com/c/siim-covid19-detection?utm_medium=email&utm_source=gamma&utm_campaign=comp-siim-covid19

4-Herramientas para el desarrollo

Se cuenta con kaggle para la codificación y ejecución y presentación en competencia.

Son cuentas individuales que luego se agrupan en equipos (kaggle cuenta general)

Y para organizar información y avance.
github, discard, trello.

5-Alcance:

A los efectos de la participación en la competencia se utilizará el entorno de desarrollo de kaggle para desarrollar la propuesta y sección de presentación. Quedando todo otro desarrollo

como por ejemplo carga de una imagen y respuesta vía web como continuación posterior a la competencia.

La competencia se trata de la presentación del código de ejecución del programa.

Las presentaciones a este concurso deben realizarse a través de Notebooks. Para que el botón "Enviar" esté activo después de una confirmación, se deben cumplir las siguientes condiciones:

- CPU portátil \leq 9 horas de tiempo de ejecución
- GPU Notebook \leq 9 horas de tiempo de ejecución
- Acceso a Internet deshabilitado
- Se permiten datos externos disponibles de forma gratuita y pública, incluidos modelos previamente entrenados
- El archivo de envío debe tener un nombre `submission.csv`

PASOS DENTRO DE ENTORNO KAGGLE

1. Primero, deberá escribir un cuaderno que lea el conjunto de datos de la competencia y haga predicciones sobre el conjunto de prueba. Específicamente, haga que su Cuaderno escriba sus predicciones en un "archivo de envío", que normalmente es un archivo de envío.csv, aunque algunas competencias tienen formatos especiales. Consulte la página de evaluación de la competencia o busque `sample_submission.csv` (o similar) en la página de datos para obtener más información sobre el nombre y el formato esperados de su archivo de envío.

2. Guarde una versión completa de su Notebook haciendo clic en "Guardar versión" y seleccionando "Guardar y ejecutar todo". Esto guarda su código, lo ejecuta y crea una versión del código y la salida. Una vez que haya terminado de guardar, vaya a la página Visor de su nueva versión de Notebook.

3. En el visor de Notebook, navegue hasta la sección Salida, busque y seleccione el archivo de envío que creó y haga clic en el botón "Enviar".

¿Cómo se determinarán los ganadores?

En algunos concursos de códigos, los ganadores se determinarán volviendo a ejecutar los blocs de notas asociados a los envíos seleccionados en un conjunto de prueba privado.

En tales concursos, creará sus modelos en Notebooks y hará envíos basados en el conjunto de prueba provisto en la página de Datos. Realizará presentaciones desde su cuaderno siguiendo los pasos anteriores y seleccionará las presentaciones para la evaluación final en la página "Mis presentaciones", de la misma manera que en una competencia normal.

Después de la fecha límite de la competencia, Kaggle volverá a ejecutar su código en un conjunto de prueba privado que no se le proporcionó. La puntuación de su modelo en comparación con este conjunto de pruebas privadas determinará su clasificación en la tabla de clasificación privada y su posición final en la competición.

¿Cómo se determinarán los ganadores?

En algunos concursos de códigos, los ganadores se determinarán volviendo a ejecutar los blocs de notas asociados a los envíos seleccionados en un conjunto de prueba privado.

En tales concursos, creará sus modelos en Notebooks y hará envíos basados en el conjunto de prueba provisto en la página de Datos. Realizará presentaciones desde su cuaderno siguiendo los pasos anteriores y seleccionará las presentaciones para la evaluación final en la página "Mis presentaciones", de la misma manera que en una competencia normal.

Después de la fecha límite de la competencia, Kaggle volverá a ejecutar su código en un conjunto de prueba privado que no se le proporcionó. La puntuación de su modelo en comparación con este conjunto de pruebas privadas determinará su clasificación en la tabla de clasificación privada y su posición final en la competición.

Hacer una presentación

Deberá enviar sus predicciones modelo para recibir una puntuación y una posición en la tabla de clasificación en una competencia. La forma de hacerlo depende del formato de la competencia.

De cualquier manera, recuerde que su equipo está limitado a una cierta cantidad de presentaciones por día. Este número es cinco, en promedio, pero varía de una competencia a otra.

Tabla de clasificación

Uno de los aspectos más importantes de las competencias de Kaggle es la tabla de clasificación:

Public Leaderboard

Private Leaderboard

The private leaderboard is calculated with approximately 75% of the test data.

This competition has completed. This leaderboard reflects the final standings.






Refresh

In the money

Gold

Silver

Bronze

#	△pub	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	Toxic Crusaders			0.9885	171	3mo
2	—	neongen & Computer says no			0.9882	129	3mo
3	▲ 3	Adversarial Autoencoder			0.9880	451	3mo
4	▲ 1	Leyantech			0.9878	164	3mo
5	▲ 2	TPMPM			0.9878	299	3mo

La tabla de clasificación de la competencia tiene dos partes.

La tabla de clasificación pública proporciona puntuaciones de envío visibles públicamente basadas en una muestra representativa de los datos enviados. Esta tabla de clasificación es visible durante toda la competencia.

La tabla de clasificación privada, por el contrario, rastrea el rendimiento del modelo en datos no vistos por los participantes. Por tanto, la tabla de clasificación privada tiene la última palabra sobre qué modelos son los mejores y, por tanto, quiénes serán los ganadores y perdedores del Concurso. El subconjunto de datos que se calcula en la tabla de clasificación privada o el rendimiento de una presentación en la tabla de clasificación privada no se divulga a los usuarios hasta que se cierra la competencia.

Muchos usuarios observan de cerca la tabla de clasificación pública, ya que los avances en la competencia se anuncian mediante ganancias de puntuación en la tabla de clasificación. Estos saltos a su vez motivan a otros equipos que trabajan en la competencia en busca de esos avances. Pero es importante mantener la tabla de clasificación pública en perspectiva. Es muy fácil sobreajustar un modelo, creando algo que funciona muy bien en la clasificación pública, pero muy mal en la privada. A esto se le llama [sobreajuste](#).

Envío de predicciones

Enviar cargando un archivo

Para la mayoría de las competencias, enviar predicciones significa cargar un conjunto de predicciones (conocido como "archivo de envío") a Kaggle.

Cualquier competencia que admita este estilo de presentación tendrá los botones "Enviar predicciones" y "Mis presentaciones" en el encabezado de la página de inicio de la competencia.

Para enviar una nueva predicción, utilice el botón Enviar predicción. Esto abrirá un modal que le permitirá cargar su archivo de envío. Intentaremos calificar este archivo y luego agregarlo a Mis envíos una vez que haya terminado de procesarse.

Tenga en cuenta que para contar, su envío primero debe pasar el procesamiento. Si su envío falla durante el paso de procesamiento, no se contará y no recibirá una puntuación; ni contará para su límite de envío

diario. Si encuentra problemas con su archivo de envío, su mejor curso de acción es pedir consejo en el foro de discusión del Concurso.

Si hace clic en la pestaña Mis envíos, verá una lista de todos los envíos que haya realizado a esta competencia. También puede utilizar esta pestaña para seleccionar qué archivo (s) de envío desea enviar para su puntuación antes de que finalice el Concurso. Su puntaje final y ubicación al final de la competencia será la presentación seleccionada que se desempeñó mejor en la tabla de clasificación privada. Si no selecciona envíos para calificar antes de que finalice la competencia, la plataforma seleccionará automáticamente aquellos que obtuvieron el mejor desempeño en la tabla de clasificación pública, a menos que se comunique lo contrario en la competencia.

Envío subiendo desde un bloc de notas

Además de nuestras competencias habituales, Kaggle también puede permitir presentaciones de competencias desde Kaggle Notebooks. Los cuadernos son un entorno de edición de código interactivo en el navegador; para obtener más información sobre ellos, consulte las secciones de documentación en [Notebooks](#).

Para construir un modelo, comience por inicializar un nuevo Notebook con el Conjunto de datos de competencia como fuente de datos. Esto se hace fácilmente yendo a la pestaña "Notebooks" dentro de la página de una competencia y luego haciendo clic en "New Notebook". El conjunto de datos de esa competencia se utilizará automáticamente como fuente de datos. Los nuevos blocs de notas serán privados de forma predeterminada, pero se pueden cambiar a públicos o compartir con usuarios individuales (por ejemplo, otros miembros de su equipo).

Cree su modelo y pruebe su rendimiento con el editor interactivo. Una vez que esté satisfecho con su modelo, utilícelo para generar un archivo de envío dentro del Cuaderno y escriba ese archivo de envío en el disco en el directorio de trabajo predeterminado (/ kaggle / working). Luego haga clic en "Guardar versión" y seleccione "Guardar y ejecutar todo" para crear una nueva versión de Notebook usando su código.

Una vez que la nueva Versión de Notebook esté lista (debe ejecutarse de arriba a abajo dentro de las restricciones de la plataforma de Notebooks), navegue a la página del Visor de Notebook para ver los resultados de la ejecución, luego busque y seleccione su archivo de envío en la sección Salida, y debería ver un botón "Enviar" para enviarlo a la competencia.

6-TEMAS PRIMERAS CHARLA Y TAREAS

- Investigar el armado de grupos en kaggle (guille)
- Alcance del proyecto (Gustavo)
- Entender como es la evaluación (David)
- Levantamiento de los datos (Franklin, David) (balanceo de clases, lectura,

Gustavo busca acerca de cómo balancear las clases en el optimizados)

7-AVANCES

hiba trabajó algo similar? gustavo

omar recomendo yolov4 en audio al grupo cuando se sugirió yolov5

ejemplo kaggle yolov4 <https://www.kaggle.com/tanreinama/yolov4-detection-sample/execution>

discord aporte franklin-

Medical X-ray \$ Image Classification using Convolutional Neural Network

Construction of CNN model for detection of pneumonia in x-rays from scratch

<https://towardsdatascience.com/medical-x-ray-%EF%B8%8F-image-classification-using-convolutional-neural-network-9a6d33b1c2a>

<https://gitlab.com/irafm-ai/poly-yolo>

<https://www.kaggle.com/alberto256/covid19-visualization-john-hopkins>

jaca hugo

Drive de Yolo4:

https://drive.google.com/drive/folders/1nDkqOiE_OzNAOu1cCbDnfBqnGxGs09Dj?usp=sharing

Gustavo Lazzeri — 25/05/2021

Hola gente. Les dejo el video de IA for medicine que habla de ajustar los pesos de cada clase cuando no están balanceadas.

Y junto con eso, algún artículo para implementarlo en TF.

IA for medicine (AI for Medical Diagnosis > Semana 1 > Impact of Class Imbalance on Loss Calculation):

<https://www.coursera.org/learn/ai-for-medical-diagnosis/lecture/qSNmX/impact-of-class-imbalance-on-loss-calculation>

Implementación (por el medio, buscar "Calculate class weights"):

https://www.tensorflow.org/tutorials/structured_data/imbalanced_data?hl=en

How does TensorFlow/Keras's class_weight parameter of the fit() function work?

<https://stackoverflow.com/questions/57934638/how-does-tensorflow-keras-class-weight-parameter-of-the-fit-function-work>

chaco

https://www.youtube.com/watch?v=Zk67BeYJY_I&list=PL1yL1TVvgf7MtnQv9LADQe-MKynJ2qm65&index=1

Gustavo

Ahí dejé en Discord material sobre ajustar un coeficiente (no digo peso que se confunde), para que las clases con menos ejemplos influyan parejo con las otras al calcular la pérdida.

Es una alternativa al data augmentation.

<https://discord.com/channels/804514441042853888/846526501967888415/846815433344286740>

(gustavo)

Probé el tema de cambiar el peso de las cada clase cuando están desbalanceadas (en un ejemplo viejo que tengo), y anda bien.

Igualmete, cada ve que entreno el modelo, varía bastante el accuracy (me da 0,8 a 0,93), así que puede ser casualidad. Pero lo voy a probar varias veces a ver si es consistente.

david

<https://wiki.cancerimagingarchive.net/display/Public/DICOM+Modality+Abbreviations>

<https://awesomeopensource.com/project/experiencor/keras-yolo3>

En este link aparece la distribución de los folders

info

Omar compartio este drive:

https://drive.google.com/drive/folders/1nDkqOiE_OzNAOu1cCbDnfBqnGxGs09Dj?usp=sharing

8-LINKS GENERALES

info general de covid

<https://covid19comisionunam.unamglobal.com/?p=84772>

video centro graduados explicación yolov4

https://www.youtube.com/watch?v=939Pd_-F_2g yolo4

Plataforma nacional investigación a futuro

<https://www.argentina.gob.ar/noticias/covid-t-una-plataforma-que-permitira-el-estudio-de-linfocitos-t-en-individuos-vacunados-y>

TAREAS A FUTURO FUERA DE ALCANCE

implementación web franklin

<https://morioh.com/p/473f7ad7a055?f=5c21fb01c16e2556b555ab32>

flask sugerencia claudio

<https://youtu.be/Esdj9wlBOal>