



ХимРар

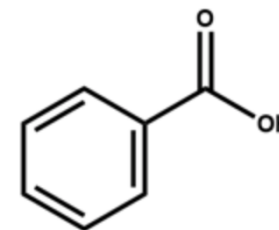
ПОСТРОЕНИЕ СХЕМЫ МОЛЕКУЛЫ

Powered by *Veggie Salad*

Идея

chemdraw-Dec-2016.cdx
ChemDraw12011615112D

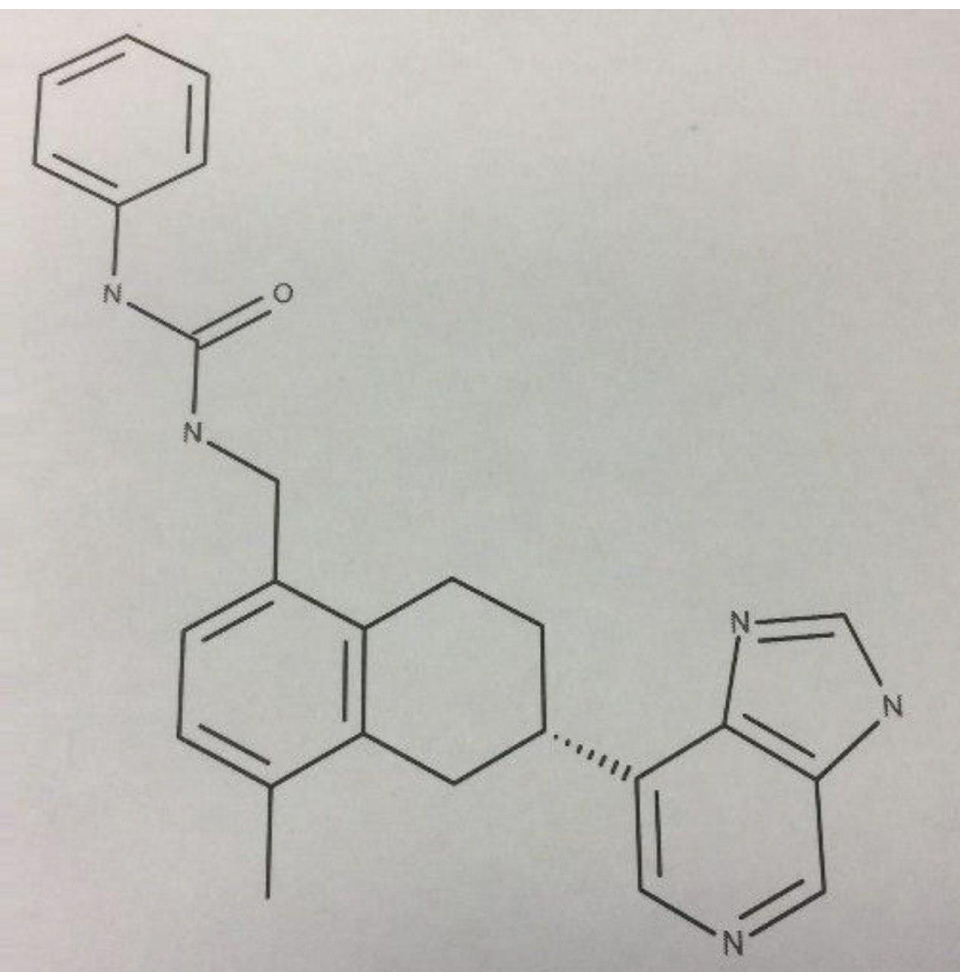
```
9 9 0 0 0 0 0 0 0 0 0999 V2000
-1.4289 -0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.4289 -0.8250 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7145 -1.2375 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 -0.8250 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 -0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7145 0.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 0.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.4289 0.0000 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 1.2375 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0
2 3 2 0
3 4 1 0
4 5 2 0
5 6 1 0
6 1 2 0
5 7 1 0
7 8 1 0
7 9 2 0
M END
```



Идея:

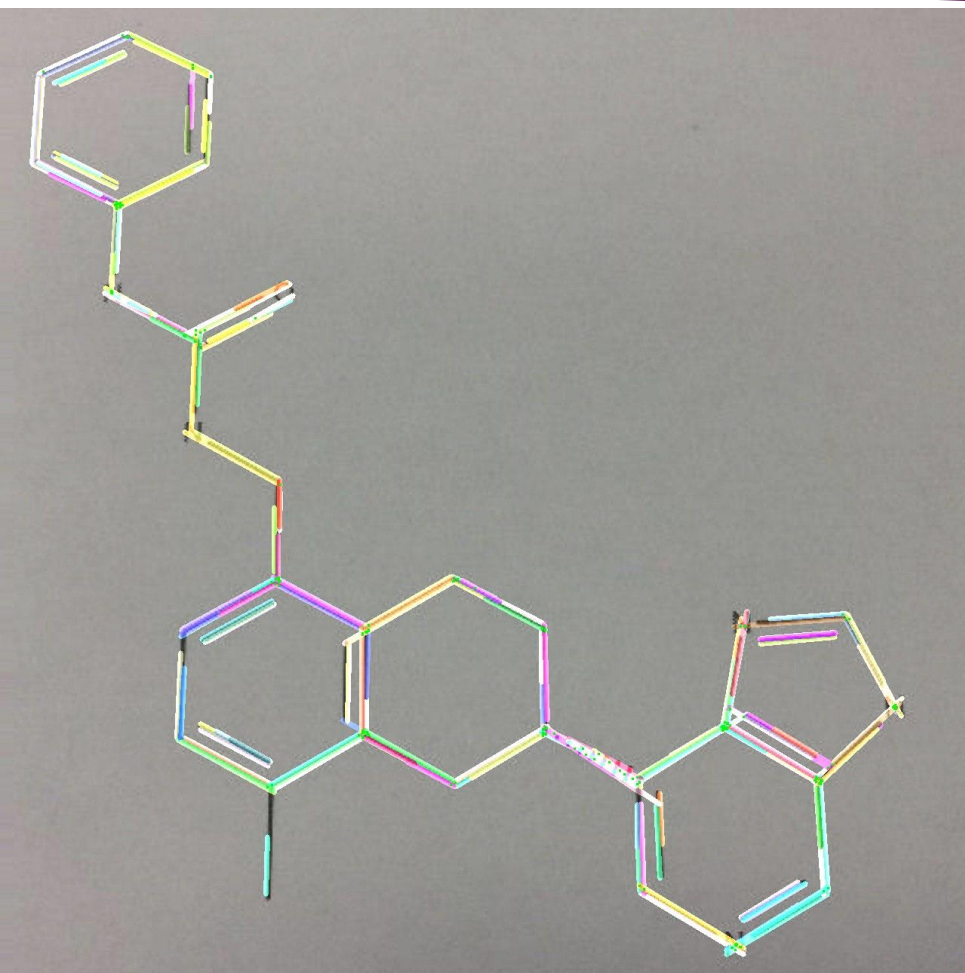
- ▶ Использовать OpenCV для определения местоположения связей (отрезков) молекулы.
- ▶ Перевод в относительные координаты узлов связей
- ▶ Использование RDkit для построения молекулы

Как это работает?



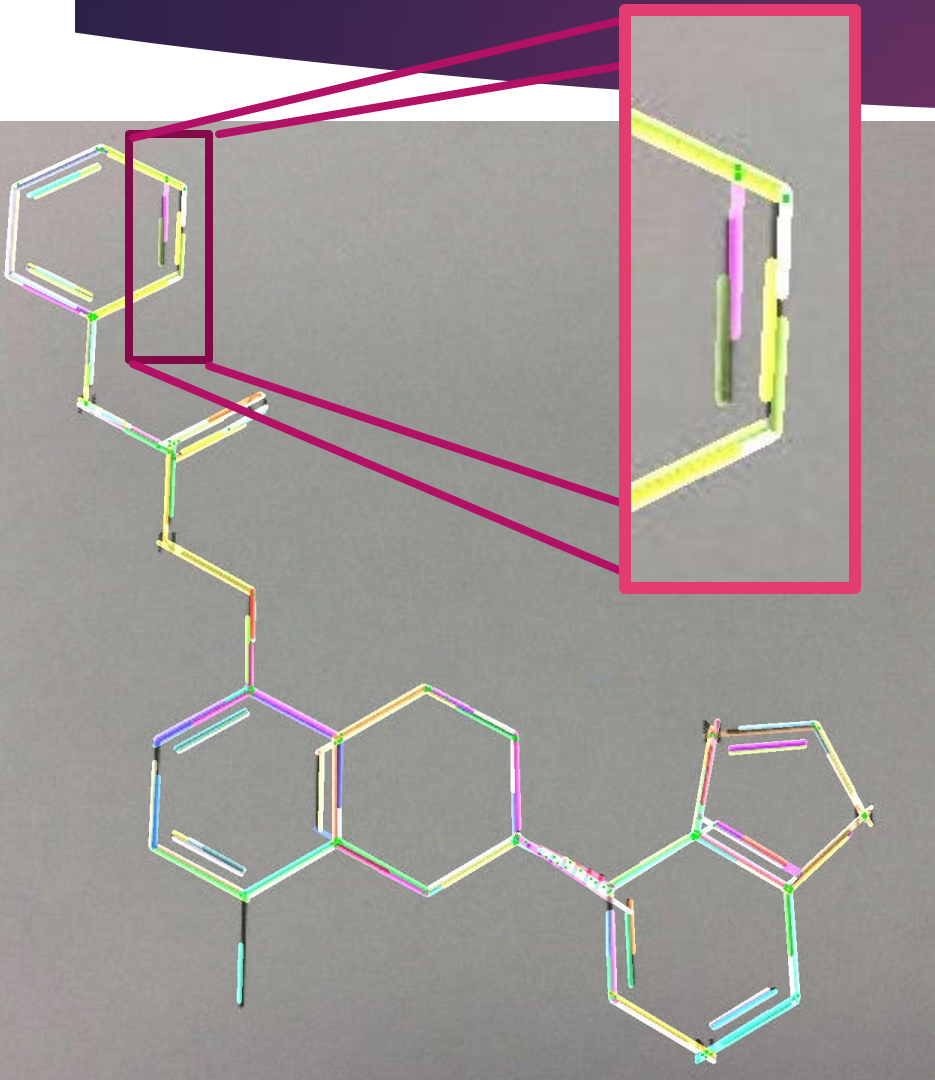
- Берем фотку молекулы

Как это работает?



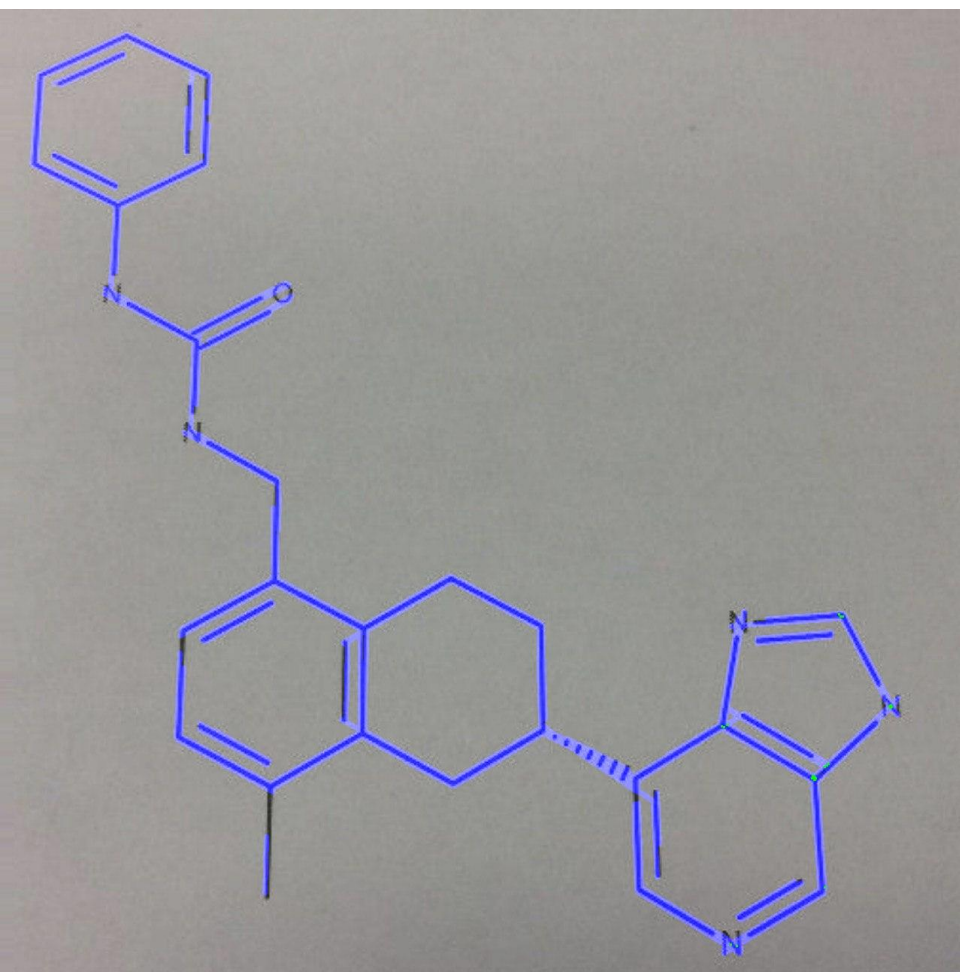
- ▶ Берем фотку молекулы
- ▶ Прогоняем через OpenCV

Проблема...



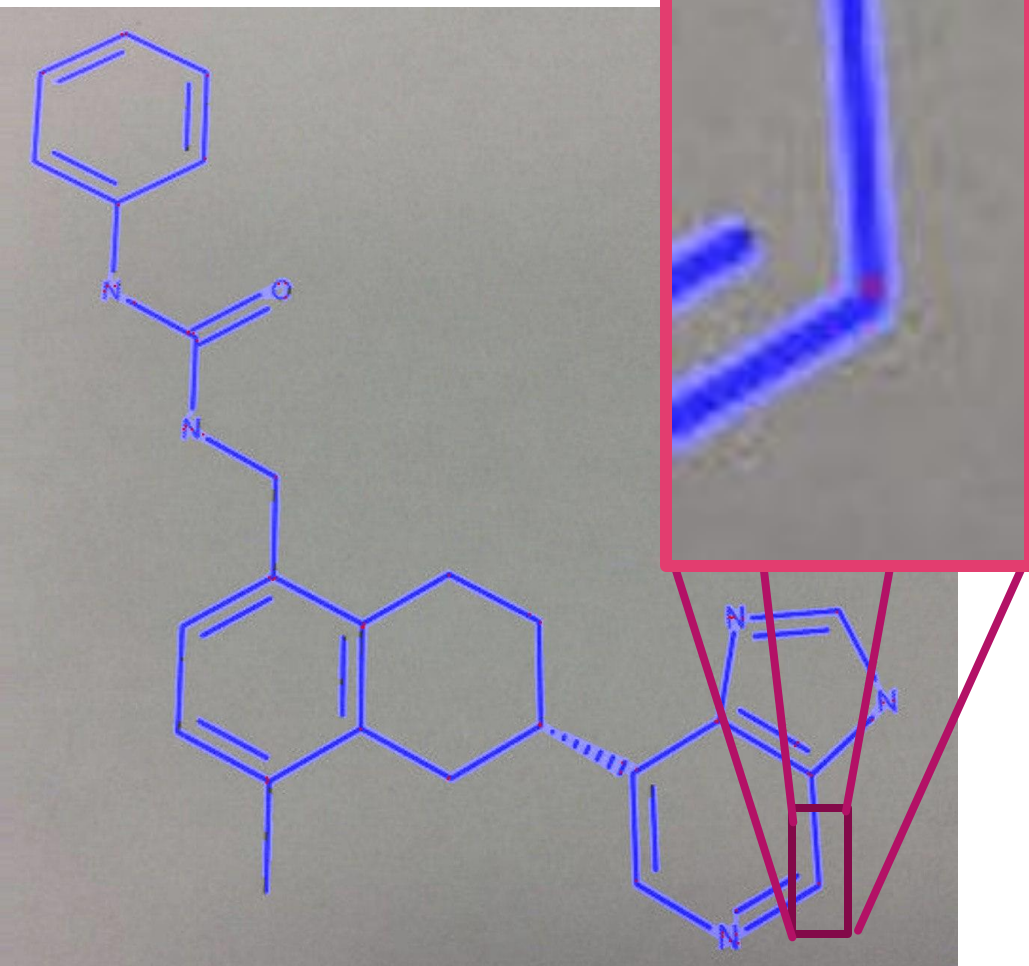
- ▶ Отрезки разбиваются на несколько...

Решение



- Посчитать среднее значение длин отрезков и передать параметры минимальной и максимальной длины отрезков в OpenCV, равные $\text{mean_value} * (1 - \text{delta1})$ и $\text{mean_value} * (1 + \text{delta2})$, соответственно. $0 < \text{delta1}, \text{delta2} < 1$.
- Прогнать OpenCV с обновленными параметрами

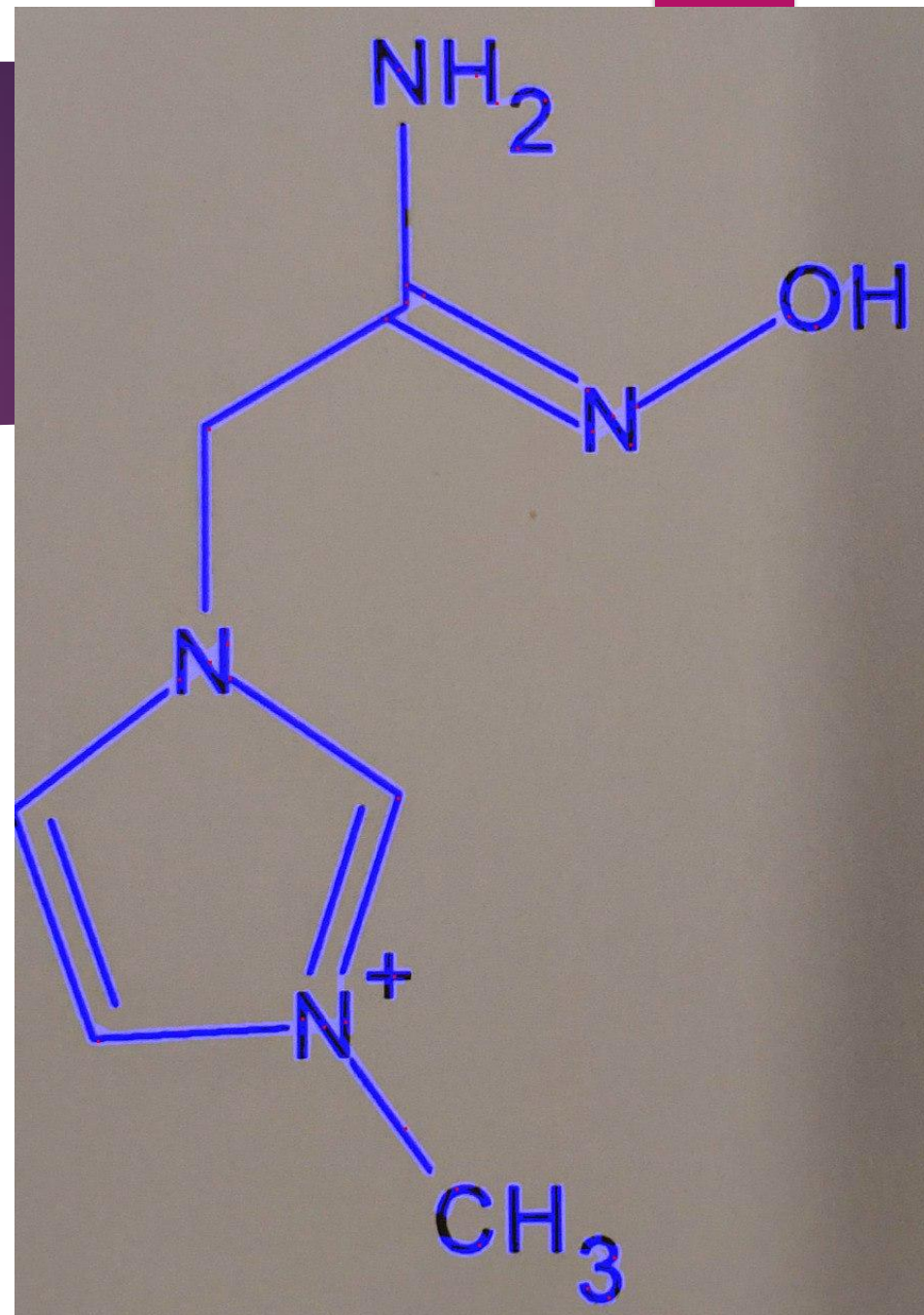
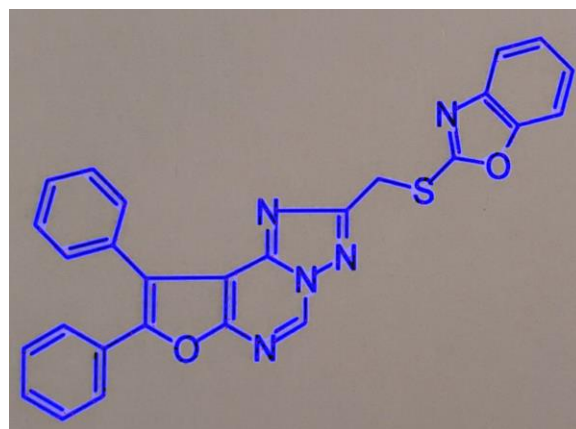
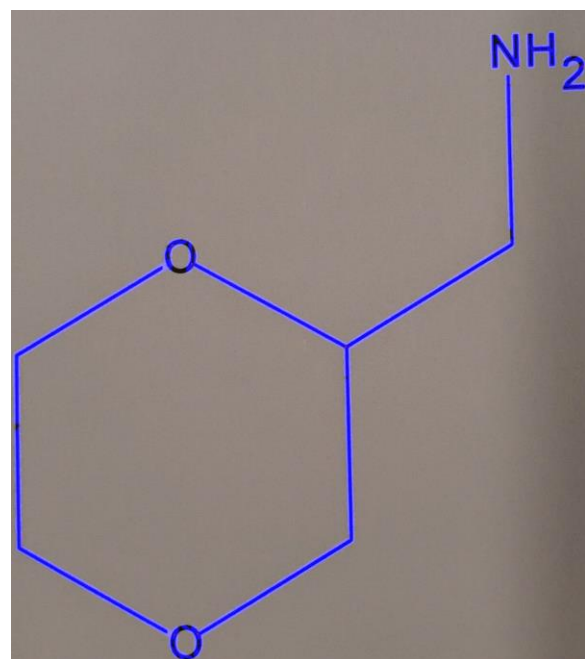
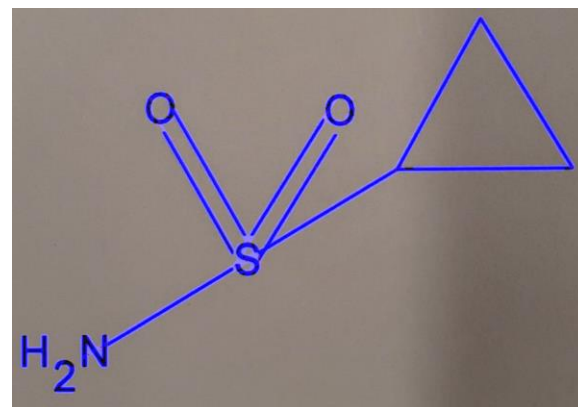
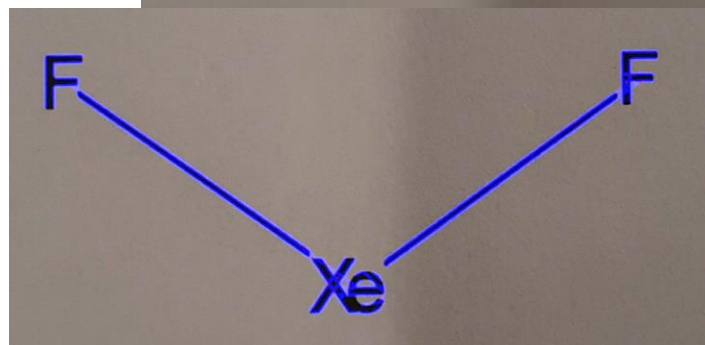
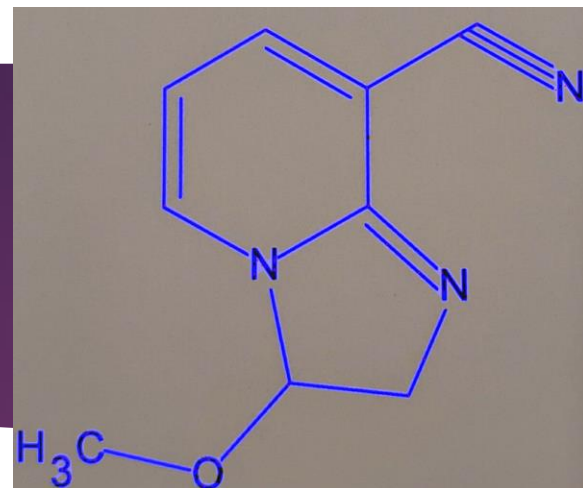
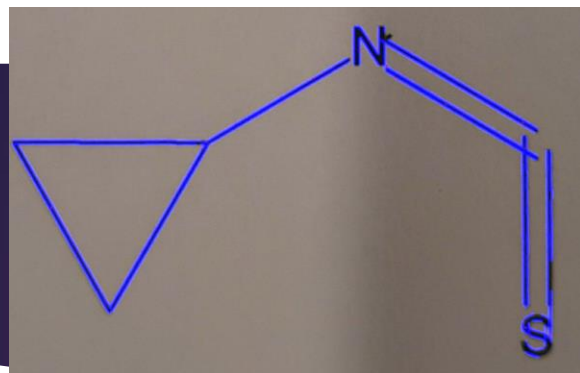
Как это работает?



- Найдем пересечения отрезков - координаты связей

Расставить химические элементы в схеме на свои места

- ▶ Идея: сегментация текста на изображении, анализ и запоминание координат их расположения.
- ▶ При построении схемы учесть эту информацию



Спасибо за внимание